2016

Propensity Score for Causal Inference of Multiple and Multivalued Treatments

Zirui Gu
Virginia Commonwealth University

PROPENSITY SCORE FOR CAUSAL INFERENCE OF MULTIPLE AND MULTIVALUED TREATMENTS

A Dissertation submitted in partial fulfillment of the requirements for the degree of DOCTOR OF PHILOSOPHY at Virginia Commonwealth University.

by

ZIRUI GU
MS, Mississippi State University, 2011
BSc, Capital Normal University, 2008

Director: BASSAM A. DAHMAN, PH.D.
AFFILIATE ASSOCIATE PROFESSOR, DEPARTMENT OF BIOSTATISTICS
ASSOCIATE PROFESSOR, DEPARTMENT OF HEALTH BEHAVIOR AND POLICY

Virginia Commonwealth University
Richmond, Virginia

November 2016

Acknowledgement

I would like to express my special appreciation and thanks to my advisor professor Dr. Bassam A. Dahman, you have been a tremendous mentor for me. I would like to thank you for encouraging my research and for allowing me to grow as a research scientist. Your advice on both research as well as on my career have been invaluable.

I would also like to thank my committee members, Professor Roy T. Sabo, Professor Nitai D. Mukhopadhyay, Dr. Sarah Hartigan, Professor Qiqi Lu, for serving as my committee members even at hardship.

A special thanks to my family. Words cannot express how grateful I am to my mother, father, and my wife for all of the sacrifices that you've made on my behalf.

Table of Contents

Page

List of Tables

Page

List of Figures

Page

Abstract

CAUSAL INFERENCE FOR MULTIPLE AND MULTIVALUED TREATMENTS

By Zirui Gu, MS

A Dissertation submitted in partial fulfillment of the requirements for the degree of Ph.D. at Virginia Commonwealth University.

Virginia Commonwealth University, 2016

Major Director: Bassam A. Dahman
Affiliate Associate Professor, Department of Biostatistics
Associate Professor, Department of Health Behavior and Policy

Propensity score methods (PSM) that have been widely used to reduce selection bias in observational studies are restricted to a binary treatment. Imai and van Dyk extended PSM to estimate non-binary treatment effect using stratification with P-Function, and generalized inverse treatment probability weighting (GIPTW). However, propensity score (PS) matching methods on multiple treatments received little attention, and existing generalized PSMs merely focused on estimates of main treatment effects but omitted potential interaction effects that are of essential interest in many studies. In this dissertation, I extend Rubin's PS matching theory to general treatment regimens under the P-Function framework. From theory to practice, I propose an innovative distance measure that can summarize similarities among subjects in multiple treatment groups. Based on this

distance measure I propose four generalized propensity score matching methodologies. The first two methods are extensions of nearest neighbor matching. I implemented Monte Carlo simulation studies to compare them with GIPTW and stratification on P-Function methods. The next two methods are extensions of the nearest neighbor caliper width matching and variable matching. I define the caliper width as the product of a weighted standard deviation of all possible pairwise distances between two treatment groups. I conduct a series of simulation studies to determine an optimal caliper width by searching the lowest mean square error of average causal interaction effect. I further compare the ones with optimal caliper width with other methods using simulations. Finally, I apply these methods to the National Medical Expenditure Survey data to examine the average causal main effect of duration and frequency of smoking as well as their interaction effect on annual medical expenditures. Using proposed methods, researchers can apply regression models with specified interaction terms to the matched data and simultaneously obtain both main and interaction effects estimate with improved statistical properties.

Chapter 1

Introduction

1.1 History of Causal Inference and Propensity Score Methods

Estimating the causal effect of treatments, interventions, and exposures on outcomes without randomization is a common objective in retrospective studies. Although observational data such as administrative records, survey data, and standardized test scores are easier to obtain than randomized controlled trial (RCT) data, the lack of treatment randomization could cause systematic difference between treatment groups on baseline confounders such as health status, age, and income levels. These sorts of discrepancies may reveal complicated correlations with the outcome, and thus lead to selection bias. Therefore, establishing accurate causal inference in non-experimental data can be more challenging than that in RCT. To adjust for confounding issues in observational data, various matching methods have been developed to replicate a randomization process as much as possible for baseline covariates. If there is no unobserved difference between treatment and control groups, and the matched pairs capture all observed difference in pretreatment variables, then matching methods could produce unbiased estimate of treatment effect. The earliest versions of matching strategies are based on two treatment groups with either a single variable or weighting several variables (Cochran and Rubin, 1973; Raynor, 1983; Rosenbaum 2002, Dehejia and Wahba 2002). When there are only few pretreatment variables, matching is easy to implement. However, as the number of covariates increases, matching become cumbersome and even infeasible. Motivated by the

high dimensionality problem, Rosenbaum and Rubin (1983), developed the propensity score method (PSM) for a paired treatment and control group. By definition, propensity score (PS) is the conditional probability of receiving treatment given a set of observed pretreatment variables. Rosenbaum and Rubin, (1983) proved under the assumption of no unobserved difference and no interference between units, controlling for propensity score instead of the large number of confounders are sufficient to estimate unbiased causal effects (Rosenbaum and Rubin 1983). Major PSMs include matching, stratification, inverse probability of treatment weighting (IPTW) and PS covariate adjustment (PSCA). Despite their popularity, the primary barrier of these PSMs is how to extend them to the scenarios of multiple or multivalued treatments. Two methodologies are developed to solve this constraint: The generalized inverse probability treatment weighting (Imbens, 2000), and stratification on P-Function (Imai and van Dyk, 2004). The generalized propensity score (GPS) is the conditional probability of receiving a particular level of treatment given the pretreatment. Imbens showed GPS was a balancing score under the assumption of weak unconfoundedness. Hereafter, several researchers used a GPS version of IPTW (Robins, Hern´an, and Brumback, 2000), or PSCA (Hirano and Imbens, 2004), to adjust for observed confounding issues. Imai and van Dyk (2004) developed the propensity function (P-Function), which is essentially a generalization of the original PS. They showed the P-Function shared analogous balancing properties of PS. They introduced a P-Function stratification method that can be applied when treatments are ordinal, categorical, continuous or multidimensional. These extensions expand the use of PSMs and have become increasingly popular. The Google Scholar citation frequency of the papers

mentioned above are 1174, 800, 240, and 318, respectively, as of Oct 29, 2013) (Zhao, van Dyk, Imai 2013). These literature, however, are restricted in two aspects. First, when the simulated or real data contains multiple treatments, they ignored the estimation of interaction effects. Interaction effect measures the direction and magnitude of an association between two variables due to a third variable. The evaluation of interaction effects is essential due to five reasons: 1) it can be useful to assess the interaction effect between two treatments or exposures. For example, in healthcare research we might be interested in whether the effect of a new hospital administrative system on patient mortality rate depends on types of department (Surgery vs. Medicine). In social science we might be interested in testing the interaction between an instructor training program and a specific course curriculum. In Epidemiology, we are often interested in interactions between one environmental exposure (i.e. Smoking, Alcohol use) and another (genetic) exposure on a disease (i.e. lung cancer, breast cancer). 2) It is useful for exploring heterogeneous treatment effects. For instance, in longitudinal studies, one primary question of interests is whether the treatment effect varies across different time point. If there are limited resources for implementing interventions in an observational study, estimation of interaction effects can help us identify subgroups that are beneficial from treatments, and subgroups that are harmful from treatments. We may want to test the interaction between a drug and patients age on the mortality rates of rare disease and find out which age group (i.e >60) benefit most from the drug. 3) In behavior sciences, the interaction effect is also called moderation effect, a moderator is a variable on which the relationship between a second variable and a dependent outcome depends. For example, the social support can be

a moderator for the relationship between stress and depression (outcome). Modeling the interaction term between stress and social support enables estimating the moderation effect of social support. 4) Leveraging interaction and main effects simultaneously can increase power in testing for the overall causal effect on an outcome (Tyler, 2014). 5) By including interaction terms, a statistical model may fit the data better. Secondly, these articles did not discuss any generalized propensity score matching methods. Matching is a traditional approach to guard against such selection bias (Cochran and Chambers 1965). An appropriate matching algorithm can classify observations into homogeneous groups so that subjects within a matched set have similar pretreatment characteristics. PS matching plays an essential role in healthcare research, economics, sociology, political science and political science (Smith 1997; Dehejia and Wahba 1999;Harding 2003; Ho et al. 2007; Armstrong, Jagolinzer, and Larcker 2009). When a treatment of interest is binary, Austin (2012) showed PS matching have the greatest potential to adjust for imbalance among observed confounders, compared to IPTW, Stratification, and PS covariate adjustment. Lu, Hornik, and Rosenbaum (2001) discussed an Optimal Non-Bipartite Matching (ONBM) method for a five-level dose treatment. The ONBM is able to capture the set of matches that minimize the sum of distances based on a given distance matrix of the overall sample. They used the propensity score calculated from McCullagh's (1980) ordinal logit model, which has been proved to be a scalar balancing score (Joffe and Rosenbaum 1999). The bipartite optimal matching is a particular case of ONBM. Gu and Rosenbaum (1993) found ONBM was not in general better than nearest neighbor matching at achieving good group balance, but did better at reducing distance within pairs. Austin

(2013) compared twelve bipartite matching algorithms. He showed nearest neighbor matching yielded the same balance in baseline covariates as did the optimal bipartite matching. Stuart (2010) suggested nearest neighbor matching is sufficient enough given the aim is to find well-matched groups rather than well-matched pairs as compared to ONBM. She further showed the nearest neighbor caliper matching had lower mean square error (MSE) than optimal bipartite matching. Austin (2012) findings indicate the ONBM do not necessarily achieve the best balance of pretreatment variables, and lowest MSE compared with other matching algorithms. In contrast, nearest neighbor caliper matching outperformed ONBM by discarding matched candidates with distance beyond a pre-specified caliper width. The rest of this chapter addresses Rubin's causal inference framework and popular PSM given one binary treatment. In section 1.2, I briefly review Rubin's causal model settings; In section 1.3, I go through the Rubin's PS theorems; In section 1.4, I review distance measures given a binary treatment. In section 1.5, I describe the nearest neighbor matching algorithm given. In section 1.6, I focus on the bipartite nearest neighbor caliper width matching algorithm. In section 1.7, I review the stratification on PS method. In section 1.8, I review the inverse probability treatment weighting method. In section 1.9, I describe common support given two treatment groups.

1.2 Causal Inference Framework Given One Binary Treatment Variable

Rubin (1974) developed the causal inference theoretical framework when the treatment is binary. Hereafter it is widely used in fields such as medical research (Christakis and Iwashyna 2003; Rubin 1997), political science (Bowers and Hansen 2005; Imai 2005), and

sociology (Smith 1997; Winship and Morgan 1999; Diprete, Thomas A. and Henriette Engelhardt. 2004; Morgan and Harding 2006;). By definition, the causal effect for an individual i is the difference between outcome $Y_i(1)$ if individual i receives the treatment and the outcome $Y_i(0)$ if individual i receives the control. (Rubin, 1974; Stuart, 2010). A major challenge of causal inference is we could only observe one potential outcome on either treatment or control for an individual. Therefore, causal inference is essentially a missing data problem. Let T_i be a treatment indicator for individual i, where $i = 1$ if the individual receives treatment, and $i = 0$ if the individual receives control. The observed outcome is: $Y_i = T_i Y_i(1) + (1 - T_i) Y_i(0)$ (Imbens, 2000; Rosenbaum, 2002). A good estimand of a causal effect is using conditional means difference between treatment and control, called average causal effect (ACE). Let τ denote the ACE, the ACE for a binary treatment is formulated as: $\tau = E[Y(1) - Y(0)]$ (Imbens, 2004). Another popular estimand, namely, average treatment effect on treated (ATT) is: $\tau_T = E[Y(1) - Y(0) \mid \mathrm{T} = 1]$ (Rubin, 1977; Imbens, 2004). This dissertation will concentrate on assessment of different PSM on estimating ACE including average causal main effect (ACME) and average causal interaction effect (ACIE).

Causal inference relies on two critical assumptions.

Assumption 1: Stable Unit Treatment Value Assumption (SUTVA) (Rubin 1980)

The outcome of one unit is assumed to be unaffected by the particular treatment assignments to the other units. If the SUTVA holds, then there is no interference between units. Thus the potential outcome of one unit is independent of another unit's treatment

assignment. This assumption may be violated in a real situation. For instance, students at a school who receive different teaching programs may interact; the visit of a patient to a hospital may depend on the previous visits; patients admitted to the same services may have interference. The causal inference is more complicated when SUTVA does not hold than the one under the SUTVA. Some researchers discussed weaker versions of SUTVA when there exist hierarchical structures in data (Hong and Raudenbush, 2006; Sobel, 2006).

Assumption 2: Unconfoundedness of Treatment Assignment (Rosenbaum, Rubin 1983b)

$$(Y(1), Y(0)) \perp T \mid X$$

The outcome is assumed to be independent of treatment assignment given observed, pretreatment variables. In other words, there is no unobserved difference within the actual treatment distribution conditioning on observed pretreatment confounders. This assumption is highly related the variable selections when estimating propensity score using a parametric model. In practice, omitting covariates can seriously bias estimates of causal effects (Rosenbaum and Rubin 1983a, Drake 1993). Hence it is extremely crucial to include all variables known to be related to both treatment assignment and the outcome in the PS model to satisfy this assumption.

Imbens (2004) proposed to test the sensitivity by estimating an effect on a variable known to be unrelated to the treatment assignment. If the test implies the effect is not significantly different from zero, then the unconfoundedness assumption is deemed to be less plausible. Heller, Rosenbaum, and Small (2009) introduced a “design sensitivity”

method to make casual effect estimate less sensitive to unobserved covariates. Methods discussed in following chapters will be strictly under these two assumptions.

1.3 Propensity Score Theorems (Rosenbaum and Rubin, 1983)

In this section, we review Rosenbaum and Rubin (1983)'s PS Theorems.

The PS is the conditional probability of receiving a treatment given pretreatment covariates, be denoted by: $\theta = pr(T = 1 | X)$. In principle, the PS summarizes the joint association between treatment assignment and pretreatment variables. They proved under the STUVA and unconfoundedness assumptions, the covariates are independent of treatment assignments conditional on the PS.

Theorem 1: PS as a balancing score (Rosenbaum and Rubin, 1983)

$$X \perp T | \theta \tag{1.3-1}$$

Theorem 1 implies assuming STUVA and unconfoundedness matching or subclassifying on the PS are equivalent to matching or stratification on X. Thus in principle, PSM could generate the same distribution in X between treatment and control groups.

Theorem 2. (Rosenbaum and Rubin, 1983)

$$(Y(1), Y(0)) \perp T | \theta \tag{1.3-2}$$

Theorem 2 indicates outcomes and treatment assignments are conditionally independent given the PS assuming STUVA and unconfoundedness. They further linked the PS and treatment assignments to estimating causal effect using Theorem 1 and 2. In section (2.1), I

will illustrate how Imai and van Dyk (2004) extended Theorem 1 and 2 to the general treatment regime.

Theorem 3. (Rosenbaum and Rubin, 1983)

Let θ be a balancing score under the unconfoundedness assumption, the ACE at θ is equal to the expected difference in observed outcome between treatment and control groups at θ

$$\mathrm{E}(Y(1) \mid \theta, T = 1) - \mathrm{E}(Y(0) \mid \theta, T = 0) = \mathrm{E}(Y(1) - Y(0) \mid \theta) \qquad (1.3\text{-}3)$$

Proof: Applying Theorem 2, we have

$$\begin{aligned} &\mathrm{E}(Y(1) \mid \theta, T = 1) - \mathrm{E}(Y(0) \mid \theta, T = 0) \\ &= \mathrm{E}(Y(1) \mid \theta) - \mathrm{E}(Y(0) \mid \theta) \\ &= \mathrm{E}(Y(1) - Y(0) \mid \theta) \end{aligned}$$

Corollary 3.1. (Rosenbaum and Rubin, 1983) Bipartite (Pair) matching on balancing scores.

Suppose there is a balancing score θ is under the unconfoundedness assumption. θ is randomly sampled from the population. Subjects from treated and control group are sampled with θ. Then the ACE at θ is equal to the expected difference in outcome between treatment and control groups for the units in the matched pairs.

Theorem 4. (Rosenbaum and Rubin, 1983) Suppose the sample size of treated group is much smaller than sample size of the control, and θ is a balancing score. The reduction in bias for any matching method using θ to match a subject from the treatment group ($T = 1$) with a subject from the control group ($T = 0$) is

$$B - B_m = \int E(X \mid \theta)\{p_m(\theta \mid T=0) - p(\theta \mid T=0)\}\,\mathrm{d}\theta \qquad (1.3\text{-}4)$$

where $B = E(X \mid T=1) - E(X \mid T=0)$, and $B_m = E(X \mid T=1) - E_m(X \mid T=0)$. B is the initial bias in X in the raw data, and B_m is the expect bias in X in the matched data. p denotes the distribution of θ in the original data. p_m denotes the distribution of θ in the matched samples. *m* represents an indicator of the matched sample distribution.

Proof:

$$\begin{aligned}
&B - B_m \\
&= E(X \mid T=1) - E(X \mid T=0) - E(X \mid T=1) + E_m(X \mid T=0) \\
&= E_m(X \mid T=0) - E(X \mid T=0) \\
&= \int \{E_m(X \mid \theta, T=0) p_m(\theta \mid T=0) - E(X \mid \theta, T=0)\, p(\theta \mid T=0)\}\,\mathrm{d}\theta \qquad (1.3\text{-}5) \\
&= \int \{E_m(X \mid \theta) p_m(\theta \mid T=0) - E(X \mid \theta)\, p(\theta \mid T=0)\}\,\mathrm{d}\theta \\
&= \int \{E(X \mid \theta) p_m(\theta \mid T=0) - E(X \mid \theta)\, p(\theta \mid T=0)\}\,\mathrm{d}\theta \\
&= \int E(X \mid \theta)(p_m(\theta \mid T=0) - p(\theta \mid T=0))\,\mathrm{d}\theta
\end{aligned}$$

The fourth equality in (1.3-5) follows from Theorem 3 (1.3-3); the fifth, from the truth that for any matching method, using PS alone to match subjects does not change the distribution of X given PS in any group of PS.

Historically, matching on PS is appealing due to three reasons: (I) Matched sets allow researchers to appreciate the similarity between treatment groups immediately. (II) Although sometimes model-based covariate adjustment is good enough to remove confounding issues, the variance of ACE is lower in matched samples than in random samples (Rubin, 1983), and the variance decreases as the difference between treatment

groups on covariates decreases (Snedecor&Cochran, 1980; Rosenbaum and Rubin, 1983). (III) Model-based adjustment on matched samples is usually more robust to departures from the assumed form of the underlying model than model-based adjustment on random samples (Rubin, 1973; 1979).

1.4 Distance Measure

Defining an appropriate distance measure is critical to quantify the similarity between two units. When matching by PS, there are two types of distance d_{ij} between subject i and subject j.

1. $d_{ij} = \left|\theta_i - \theta_j\right|$.
2. $d_{ij} = \left|\text{logit}(\theta_i) - \text{logit}(\theta_j)\right|$, where θ_k is the PS for the individual k.

Researchers found the logit of PS is closed to normal distribution, and matching on the logit of PS is better in bias reduction than directly matching on PS (Rosenbaum and Rubin,1985b; Rubin and Thomas,1996; Rubin, 2001). Other distance measures include Mahalanobis distance (Prasanta Chandra, 1936): $d_{ij} = \sqrt{(X_i - X_j)' S^{-1} (X_i - X_j)}$, where S is the variance-covariance matrix of X in the pooled treated and control sample. Exact matching requires the covariate between two units is the same. Thus it can easily end up with discarding many subjects and increased bias as compared to other matching algorithms that do not require exact match (Rosenbaum and Rubin, 1985b). Matching on Mahalanobis distance does not perform well when the dimension of covariates is high, or they are not normally distributed (Gu and Rosenbaum, 1993). In chapter 2, I will introduce

a new distance measure based on a vector of parameter estimate that uniquely indexes the P-Function, a generalized version of the propensity score.

1.5 Bipartite 1:1 Nearest Neighbor Matching (Rubin, 1973a)

Bipartite nearest neighbor matching or greedy matching method is one of the most popular and straightforward PSM. The word 'Bipartite' indicates there is a pair of treatment groups. It is the most efficient approach for one binary treatment variable setting (Stuart, 2010). The 1:1 bipartite nearest neighbor matching method starts with randomly selecting a treated subject i and finding a matched control subject with the minimum distance from the subject i. The second step includes two alternatives: match with replacement or without replacement. Match with replacement allows each unit from the control group to match multiple times. In contrast, matching without replacement limits each untreated unit to match for only once. In a binary treatment setting, the choice between match with or without replacement is associated with the trade-off between bias and variance. Match with replacement always forms a matched set with the closest distance and hence results in smaller bias than matching without replacement. If there are much less control than treated units, match with replacement is in favor of match without replacement (Dehejia and Wahba, 2002). Furthermore, the order of choosing treated units in the matching process does not affect the results for matching with replacement. On the other hand, matching without replacement may not find a control as close as the one for matching with replacement. Matching without replacement increased bias but could

improve the precision of causal effect estimates (Dehejia and Wahba, 2002). The last step is to repeat step 1 and step 2 until all treated subjects find their matched control subjects.

1.6 Bipartite Nearest Neighbor Caliper Width Matching Without Replacement (Rosenbaum and Rubin, 1985; Austin, 2013)

The Bipartite Nearest Neighbor Caliper Width Matching inherits the advantages of both caliper matching and nearest neighbor matching. The caliper width matching (Cochran and Rubin, 1973) define a caliper as: where w is an artificial weight, S_1^2 is the sample variance of the propensity score in the treatment group, and S_0^2 is the sample variance of the propensity score in the control group. With this caliper width, one could find a subset of controls and select the best candidate using the nearest neighbor matching. Cochran and Rubin (1973) found when $w = 0.2$, over 98% bias of ACE was removed. Austin (2013) found nearest neighbor caliper width with $w = 0.2$ tended to result in lowest mean squared error (MSE) among 12 PS matching methods. I will provide the algorithm details below:

1. Randomly order treated units.
2. Perform caliper matching using propensity score.
3. Conduct Nearest neighbor matching within caliper width.
4. Remove treated individuals and matched controls from the lists of treated and control units.
5. Repeat step 2 to 4 until all units from the treatment groups has been used.

1.7 Stratification on PS (Rosenbaum and Rubin, 1983b)

Stratification on PS is one of the most popular PSM for causal inference. The general idea is one can equally subclassify samples into mutually exclusive stratum using the estimated PS. Within each subset, we could either take the expected difference in responses between treated and control groups or estimate ACE using a regression model within each stratum. Within each PS stratum, treated and untreated units are roughly similar in the distribution of PS. Therefore, if the PS is correctly specified, subjects from two treatment groups are approximately similar within the same subset. Next, we can obtain the overall ACE by pooling stratum-specific estimates if ACE across stratum. Rosenbaum and Rubin (1984) showed 5 strata could eliminate over 90% of the bias of ACE estimate. Hullsiek and Louis (2002) addressed that increasing number of strata could increase bias reduction but may be at an expense of increase in variance.

1.8 Inverse Probability Treatment Weighting (IPTW)

Another commonly used PSM for estimating ACE is Inverse Probability Treatment Weighting (IPTW) (Rosenbaum, 1987; Imbens, 2000). The IPTW is motivated by two equations: $E\left(\frac{TY}{\theta}\right) = E\left(Y(1)\right)$, and $E\left(\frac{(1-T)Y}{\theta}\right) = E\left(Y(0)\right)$, where θ denotes the PS, T is the treatment indicator, Y represents the response. Because these two equation hold, the IPTW estimate of ACE is: $\sum_{i=1}^{n}\left(\frac{T_i Y_i}{\theta} - \frac{(1-T_i)Y_i}{1-\theta}\right)$, where n denotes the sample size. Robins, Hern´an, and Brumback (2000) found this type of estimate could be unstable

because the PS may vary considerably with individual covariates. To solve this issue, they proposed to replace the inverse of PS by $\frac{p_{\varphi}(T_i)}{\theta}$, where $p_{\varphi}(T_i)$ denotes a parametric model for the marginal distribution of T. They proved this modified estimator could substantially mitigate the instability. Hirano, Imbens and Ridder (2003) provided another improvement solution as $\frac{1}{\theta}\sum_{i=1}^{n}\frac{T_i}{\theta}$, which showed similar performance to $\frac{p_{\varphi}(T_i)}{\theta}$.

1.9 Common Support with a Binary Treatment

Common support is the overlapped region in the distribution of PS between treatment and control groups. In practice, checking common support is a key step to determine whether it is reliable to estimate ACE. Matching within common support region excludes units with extreme values of from a treatment group that is far away from other groups, and thus help improve the quality the matched set. Flores (2009) showed in using observations from treatment group in locations with indigent overlap in the distribution of generalized propensity scores could yield more biased ACE estimate. He further showed using treatment groups with better common support could reduce bias. If most of the controls are outside the range of treated units, then any PSMs may not be efficient to estimate ACE without extensive extrapolation.

This dissertation is structured as follows: In chapter 2, I review the P-Function theorems (Imai and van Dyk, 2004). Based on the P-Function framework, I extend Rubin's PS matching theorems (Rubin and Rosenbaum, 1983) given general treatment regime. In

chapter 3, I propose an innovative distance measure that summarize the similarity structure of multiple treatment groups. Based on this distance measure, I extend the bipartite nearest neighbor matching with and without replacement to multiple and multivalued treatment scenarios. I conduct Monte Carlo simulations to compare the performance of these methods with other two popular methods: generalized inverse probability treatment weighing (GIPTW) (Imbens, 2000) and stratification on P-Function (SP) (Imai and van Dyk, 2004). In chapter 4, I introduce an innovative caliper width and generalize the bipartite nearest neighbor caliper width matching (Dehejia and Wahba 2002, Austin 2010) to the non-bipartite version. I implement a series of simulations to evaluate the optimal caliper width under the various prevalence of treatment settings. I further compare each optimal caliper width methods with methods discussed in chapter 3. I further apply these methods to the National Medical Expenditure Survey data to examine the average causal effects of duration and frequency of smoke and their interaction effect. Finally, we review the main contributions of this research and discuss limitations and possible extensions in chapter 5.

Chapter 2

Generalized Causal Inference under the P-Function Framework

In this chapter, I first go through the P-Function theorems (Imai and van Dyk, 2004). Then I extend Rubin's PS matching theorems (Rubin and Rosenbaum, 1983) to general treatment regime using the P-Function Framework.

2.1 P-Function (Imai and van Dyk, 2004)

I review the P-Function notations, assumptions, and Theorems in this section. Assume a random sample includes n individuals. For an individual $i = 1,...,n$, there is a $p \times 1$ vector of pretreatment variables $\mathbf{X}_i$. Each receives a possibly multi-dimensioned treatment, intervention, or exposure $\mathbf{T}_i$. The outcome value corresponding to $\mathbf{T}_i$ is $Y_i(\mathbf{t})$, for $i = 1,...,\mathrm{n}$, $\mathbf{t}$ is a particular vector of treatment levels mapping to $Y_i(\mathbf{t})$, $\mathbf{t} \in \Gamma$. $Y_i(\mathbf{t})$ and $\mathbf{T}_i$ are random variables. Note here boldface represents multivariate variables, non-bold is for univariate variables. The P-Function is the conditional probability density function of the treatment given observed pretreatment covariates. Ψ accounts for a vector of parameters that parameterize the P-Function. The generalized causal inference requires three key assumptions.

Assumption 1: STUVA (Rubin 1980). The potential outcomes for one subject is independent of potential treatment status of another subject given observed covariates.

Assumption 2: Unconfoundedness (Rubin and Rosenbaum 1983). The distribution of treatment and potential outcomes are independent of each other given observed covariates. The first two assumptions are essentially extended from the STUVA and unconfoundedness assumption addressed in section 1.2 (Page 9-10). The only difference is P-Function relax the distribution of $\mathbf{T}$ to any forms. Formally, we have $Y(\mathbf{t}) \perp \mathbf{T} \mid \mathbf{X}$ corresponded to the generalised unconfoundedness assumption.

Under this assumption, the outcomes do not depend on treatment assignments given pretreatment confounders. Thus we have $Y(\mathbf{t}) \perp \mathbf{T} \mid \mathbf{X} \Leftrightarrow p\{\mathbf{T} \mid Y(\mathbf{t}), \mathbf{X}\} = p\{\mathbf{T} \mid \mathbf{X}\}$. This equation implies we could model $\mathbf{T}$ conditioning on $\mathbf{X}$ without considering $Y(\mathbf{t})$.

Assumption 3: Uniquely Parameterized P-Function (UPPF)

For each $\mathbf{X}$, there exists a unique finite dimensional parameter, $\theta \in \Theta$, such that $p_\psi(\mathbf{T} \mid \mathbf{X})$ depends on $\mathbf{X}$ only through $\theta_\psi(\mathbf{X})$. Under this assumption, we may rewrite $p_\psi(\mathbf{T} \mid \mathbf{X})$ as $e\{. \mid \theta\}$, because θ uniquely characterizes $p_\psi(\mathbf{T} \mid \mathbf{X})$. Therefore, rather than doing matching or stratification on $\mathbf{X}$, we can do matching or stratification on θ, which usually has much lower dimension than $\mathbf{X}$. Based on the nature of $\mathbf{T}$, $p_\psi(\mathbf{T} \mid \mathbf{X})$ can be estimated using a parametric model with a set of parameters ψ. For example, in a traditional setting, T is binary, $\Gamma = \{0,1\}$. We may assume the P-Function is a Bernoulli density function. $T \mid \mathbf{X} \sim Bernoulli(\theta)$, where θ is the true propensity score. We may estimate θ using a logistic regression model and ψ is a vector of coefficients for the logit model. If there are two binary treatments, then $\Gamma = \{(1,1),(1,0),(0,1),(0,0)\}$. We may

assume the P-Function follows a bivariate Bernoulli density, $\boldsymbol{T} \mid \mathbf{X} \sim Bernoulli(\theta_1, \theta_2)$, where θ_1, θ_2 are two propensity scores that uniquely index the P-Function. We may model $p_\psi(\mathbf{T} \mid \mathbf{X})$ by employing a bivariate logistic regression model. If there are two categorical treatments, we can fit a bivariate multinomial logit model (Engel, J.,1988). θ in this setting is a vector of two generalized propensity scores: the conditional probability of receiving a treatment that the individual received. If there are two continuous treatments, we may assume the P-Function follows a bivariate normal distribution, $\boldsymbol{T} \mid \mathbf{X} \sim N(\mathbf{X}'\boldsymbol{\beta}, \mathbf{V})$, where $\theta = \mathbf{X}'\boldsymbol{\beta}$ uniquely represents $p_\psi(\mathbf{T} \mid \mathbf{X})$, and $\psi = (\boldsymbol{\beta}, \mathbf{V})$. The UPPF assumption facilitates the process of matching or stratification on P-Function.

Theorem 5 (Imai and van Dyk, 2004): The treatment distribution is independent of pretreatment covariates given the P-Function:

$$\mathbf{T} \perp \mathbf{X} \mid e(. \mid \boldsymbol{\theta}) \qquad (2.1\text{-}1)$$

Proof: Based on the UPPF assumption and definition of θ, θ is sufficient for the conditional distribution: $p(\mathrm{T} \mid \mathbf{X})$, thus we have for $e(. \mid \boldsymbol{\theta}) = e(. \mid \mathbf{X})$,

$$p(\mathbf{T} \mid e(. \mid \mathrm{X})) = p(\mathbf{T} \mid e(. \mid \boldsymbol{\theta})) = p(\mathbf{T} \mid \boldsymbol{\theta}) = p(\mathbf{T} \mid \boldsymbol{\theta}(\mathrm{X})) = p(\mathrm{T} \mid \mathbf{X})$$

Further based on (2.2) $e(. \mid \boldsymbol{\theta})$ is redundant given $\mathbf{X}$, we have

$$p(\mathbf{T} \mid \mathbf{X}) = p(\mathbf{T} \mid e(. \mid \boldsymbol{\theta}), \mathbf{X}) = p(\mathbf{T} \mid e(. \mid \boldsymbol{\theta}))$$

(2.3) indicates $p(\mathbf{T} \mid e(. \mid \boldsymbol{\theta}), \mathbf{X})$ do not depend on $\mathbf{X}$. Therefore $\mathbf{T} \perp \mathbf{X} \mid e(. \mid \boldsymbol{\theta})$, the P-Function is a balancing score as well as θ.

Theorem 6 (Imai and van Dyk, 2004): Unconfoundedness Given P-Function

$$Y(\mathbf{t}) \perp \mathbf{T} \mid e\{. \mid \boldsymbol{\theta}\} \tag{2.1-2}$$

Proof: I will apply the law of total expectation, assumption 1 and 2 to show

$p(\mathbf{T} = \mathbf{t} \mid Y(\mathbf{t}), e(. \mid \boldsymbol{\theta})) = p(\mathbf{T} = \mathbf{t} \mid e(. \mid \boldsymbol{\theta})) = e(. \mid \boldsymbol{\theta})$, which implies (2.4)

$$\begin{aligned}
&p(\mathbf{T} = \mathbf{t} \mid Y(\mathbf{t}), e(. \mid \boldsymbol{\theta})) = \mathrm{E}[\mathbf{T} = \mathbf{t} \mid Y(\mathbf{t}), e(. \mid \boldsymbol{\theta})] \\
&= \mathrm{E}[\mathrm{E}[\mathbf{T} \mid Y(\mathbf{t}), e(. \mid \boldsymbol{\theta})] \mid Y(\mathbf{t}), e(. \mid \boldsymbol{\theta})] \\
&= \mathrm{E}[\mathrm{E}(\mathbf{T} \mid Y(\mathbf{t}), e(. \mid \boldsymbol{\theta}), \mathbf{X}) \mid Y(\mathbf{t}), e(. \mid \boldsymbol{\theta})] \\
&= \mathrm{E}[\mathrm{E}(\mathbf{T} \mid Y(\mathbf{t}), \mathbf{X}) \mid Y(\mathbf{t}), e(. \mid \boldsymbol{\theta})] \\
&= \mathrm{E}[\mathrm{E}(\mathbf{T} \mid \mathbf{X}) \mid Y(\mathbf{t}), e(. \mid \boldsymbol{\theta})] \\
&= \mathrm{E}[\mathrm{p}(\mathbf{T} \mid \mathbf{X}) \mid Y(\mathbf{t}), e(. \mid \boldsymbol{\theta})] \\
&= \mathrm{E}[e(. \mid \boldsymbol{\theta}) \mid Y(\mathbf{t}), e(. \mid \boldsymbol{\theta})] \\
&= e(. \mid \boldsymbol{\theta})
\end{aligned}$$

The second equality is by the law of total expectation. The third and fourth equality follow from (2.1) and Assumption 1. The fifth equality follows from Assumption 2.

Let $p(\mathbf{T}, Y(\mathbf{t}), \mathbf{X} \mid e(. \mid \boldsymbol{\theta}))$ be the conditional joint probability density function of $\mathbf{T}$, $Y(\mathbf{t})$, and $\mathbf{X}$ given $e(. \mid \boldsymbol{\theta})$.

2.2 P-Function Matching Theory

In this section, I propose two theorems and two corollary under the P-Function framework. These theorems are extended from the PS Theorems (Rubin, 1983b) addressed in section 1.3.

Suppose there are k disjoint treatment groups: $\mathbf{t}_1,...,\mathbf{t}_k$. The sample size for $\mathbf{t}_i$ is n_i, $i=1,...,k$. $\mathbf{t}_1,...,\mathbf{t}_k$ is ordered by ascending sample size, $n_1 \le n_2,...,\le n_{i-1} \le n_i,....,\le n_{k-1} \le n_k$, and $\mathbf{t}_i \cap \mathbf{t}_j = \varnothing, i \ne j$. The initial bias in X between t_i and t_j is

$$B_{ij} = E(\mathrm{X} \mid \mathbf{t} = \mathbf{t}_i) - E(\mathrm{X} \mid \mathbf{t} = \mathbf{t}_j) \tag{2.2-1}$$

Since there are k treatment groups, we need to consider the bias in X between any two groups. Hence the total initial bias in $\mathbf{X}$ between all treatment groups is

$$B = \sum_{i=1}^{k-1} \sum_{j=i+1}^{k} B_{ij} = \sum_{i=1}^{k-1} \sum_{j=i+1}^{k} E(\mathrm{X} \mid \mathbf{t} = \mathbf{t}_i) - E(\mathrm{X} \mid \mathbf{t} = \mathbf{t}_j) \tag{2.2-2}$$

Suppose we match $\mathbf{t}_1$ to other k-1groups, the total bias in the matched sample is

$$B_m = \sum_{i=1}^{k-1} \sum_{j=i+1}^{k} B_{mij} = \sum_{i=1}^{k-1} \sum_{j=i+1}^{k} E_m(\mathrm{X} \mid \mathbf{t} = \mathbf{t}_i) - E_m(\mathrm{X} \mid \mathbf{t} = \mathbf{t}_j) \tag{2.2-3}$$

The subscript m in (2.2-3) indicates the distribution in matched samples. If $B_m = \lambda B$ for some scalar λ, with $0 \le \lambda < 1$, then percent bias reduction in each coordinate of X is $100(1-\lambda)\%$. If $\lambda = 0$, the matched treatment groups are the same in the distribution of X.

Theorem 7: Let $\boldsymbol{\theta} = \boldsymbol{\theta}_\psi(\mathbf{X})$, be a vector of parameters that uniquely indexes a P-Function $e(.\mid \boldsymbol{\theta})$. For any matching method that uses $\boldsymbol{\theta}_\psi(\mathbf{X})$ alone that match each subject from $\mathbf{t} = \mathbf{t}_1$, with one subject from each of other $k-1$ groups: $\mathbf{t}_2,...,\mathbf{t}_k$, the reduction in bias is

$$B - B_m = \sum_{i=1}^{k-1} \sum_{j=i+1}^{k} \int E(\mathrm{X} \mid \boldsymbol{\theta}) \{ p(\boldsymbol{\theta} \mid \mathbf{t} = \mathbf{t}_i) - p(\boldsymbol{\theta} \mid \mathbf{t} = \mathbf{t}_j) + p_m(\boldsymbol{\theta} \mid \mathbf{t} = \mathbf{t}_j) - p_m(\boldsymbol{\theta} \mid \mathbf{t} = \mathbf{t}_i) \} d\boldsymbol{\theta} \tag{2.2-4}$$

where $\mathrm{p}_m(\boldsymbol{\theta}|\mathbf{t}=\mathbf{t}_i)$ denotes the distribution of $\boldsymbol{\theta}$ in the matched samples from the $\mathbf{t}_i$ group

Proof. Take the difference between (2.2-2) and (2.2-3), we have

$$
\begin{aligned}
B - B_m &= \sum_{i=1}^{k-1}\sum_{j=i+1}^{k}\{E(\mathrm{X}\,|\,\mathbf{t}=\mathbf{t}_i) - E(\mathrm{X}\,|\,\mathbf{t}=\mathbf{t}_j) + E_m(\mathrm{X}\,|\,\mathbf{t}=\mathbf{t}_j) - E_m(\mathrm{X}\,|\,\mathbf{t}=\mathbf{t}_i)\} \\
&= \sum_{i=1}^{k-1}\sum_{j=i+1}^{k}\int\{(E(\mathrm{X}\,|\,\mathbf{t}=\mathbf{t}_i,\boldsymbol{\theta})\,p(\boldsymbol{\theta}|\mathbf{t}=\mathbf{t}_i) - E(\mathrm{X}\,|\,\mathbf{t}=\mathbf{t}_j,\boldsymbol{\theta})\,p(\boldsymbol{\theta}|\mathbf{t}=\mathbf{t}_j) \\
&\quad + E_m(\mathrm{X}\,|\,\mathbf{t}=\mathbf{t}_j,\boldsymbol{\theta})\,p_m(\boldsymbol{\theta}|\mathbf{t}=\mathbf{t}_j) - E_m(\mathrm{X}\,|\,\mathbf{t}=\mathbf{t}_i,\boldsymbol{\theta})\,p_m(\boldsymbol{\theta}|\mathbf{t}=\mathbf{t}_i)\}d\boldsymbol{\theta}
\end{aligned}
\tag{2.2-5}
$$

Where the second equality in (2.2-5) follows from (2.2-6)

$$
\begin{aligned}
&\int E(\mathrm{X}\,|\,\mathbf{t}=\mathbf{t}_l,\boldsymbol{\theta})\,p(\boldsymbol{\theta}|\mathbf{t}=\mathbf{t}_l)d\boldsymbol{\theta} \\
&= \iint \mathrm{X}\,p(\mathrm{X}\,|\,\mathbf{t}=\mathbf{t}_l,\boldsymbol{\theta})p(\boldsymbol{\theta}|\mathbf{t}=\mathbf{t}_l)d\,\mathrm{X}\,d\boldsymbol{\theta} \\
&= \iint \mathrm{X}\,\frac{p(\mathrm{X},\mathbf{t}=\mathbf{t}_l,\boldsymbol{\theta})}{p(\mathbf{t}=\mathbf{t}_l,\boldsymbol{\theta})}\,\frac{p(\boldsymbol{\theta},\mathbf{t}=\mathbf{t}_l)}{p(\mathbf{t}=\mathbf{t}_l)}\,d\,\mathrm{X}\,d\boldsymbol{\theta} \\
&= \iint \mathrm{X}\,\frac{p(\mathrm{X},\mathbf{t}=\mathbf{t}_l,\boldsymbol{\theta})}{p(\mathbf{t}=\mathbf{t}_l)}d\,\mathrm{X}\,d\boldsymbol{\theta} \\
&= \iint \mathrm{X}\,p(\mathrm{X},\boldsymbol{\theta}|\mathbf{t}=\mathbf{t}_l)d\,\mathrm{X}\,d\boldsymbol{\theta} \\
&= \int \mathrm{X}\,d\,\mathrm{X}\int p(\mathrm{X},\boldsymbol{\theta}|\mathbf{t}=\mathbf{t}_l)d\boldsymbol{\theta} \\
&= \int \mathrm{X}\,p(\mathrm{X}|\mathbf{t}=\mathbf{t}_l)d\,\mathrm{X} \\
&= E(\mathrm{X}\,|\,\mathbf{t}=\mathbf{t}_l)
\end{aligned}
\tag{2.2-6}
$$

Because for any matching method, using $\boldsymbol{\theta}$ alone to match subjects does not change the distribution of X given $\boldsymbol{\theta}$ in any group $\mathbf{t}_i$, $i=1,...,k$, hence we have

$$E(\mathrm{X}\,|\,\mathbf{t}=\mathbf{t}_i,\boldsymbol{\theta}) = E_m(\mathrm{X}\,|\,\mathbf{t}=\mathbf{t}_i,\boldsymbol{\theta}) \tag{2.2-7}$$

However, by Theorem 4 (2.1-1)

$$E(\mathrm{X}\,|\,\mathbf{t}=\mathbf{t}_i,\boldsymbol{\theta}) = E(\mathrm{X}\,|\,\mathbf{t}=\mathbf{t}_j,\boldsymbol{\theta}) = E(\mathrm{X}\,|\,\boldsymbol{\theta}) \tag{2.2-8}$$

Applying (2.2-7) and (2.2-8) to (2.2-5) yields the result (2.2-4)

Theorem 7 generalizes Theorem 3 (Rosenbaum and Rubin, 1983) in section 1.3. Theorem 3 could apply to one binary treatment, and it assumes each unit from the treatment group is used for matching. Theorem 7 applies to any form of treatment variables. Moreover, it assumes not all treatment units are necessarily used for matching. Formally, it is possible that $E(\mathrm{X} \mid \mathbf{t} = \mathbf{t}_1) \neq E_m(\mathrm{X} \mid \mathbf{t} = \mathbf{t}_1)$. For example, if we perform a caliper width matching, it is possible that one subject cannot find any matched units from other groups given a relative small caliper width.

Corollary 7.1: If $E(\mathrm{X} \mid \boldsymbol{\theta}) = \alpha + \beta f(\boldsymbol{\theta})$ for some vectors α and β and some scalar-valued function $f(.)$, then matching on $\boldsymbol{\theta}$ alone equals the reduction in percent bias between treatment groups.

Proof. Apply $E(\mathrm{X} \mid \boldsymbol{\theta}) = \alpha + \beta f(\boldsymbol{\theta})$ to (2.2-4), the reduction in percent bias (RPB) for the p th coordinates of X is

$$
\begin{aligned}
RPB &= 100 \frac{\sum_{i=1}^{k-1} \sum_{j=i+1}^{k} \{E(\mathrm{X} \mid \mathbf{t} = \mathbf{t}_i) - E(\mathrm{X} \mid \mathbf{t} = \mathbf{t}_j) + E_m(\mathrm{X} \mid \mathbf{t} = \mathbf{t}_j) - E_m(\mathrm{X} \mid \mathbf{t} = \mathbf{t}_i)\}}{\sum_{i=1}^{k-1} \sum_{j=i+1}^{k} \{E(\mathrm{X} \mid \mathbf{t} = \mathbf{t}_i) - E(\mathrm{X} \mid \mathbf{t} = \mathbf{t}_j)\}} \\
&= 100 \frac{\sum_{i=1}^{k-1} \sum_{j=i+1}^{k} \int (\alpha_p + \beta_p f(\boldsymbol{\theta}))\{p(\boldsymbol{\theta} \mid \mathbf{t} = \mathbf{t}_i) - p(\boldsymbol{\theta} \mid \mathbf{t} = \mathbf{t}_j) + p_m(\boldsymbol{\theta} \mid \mathbf{t} = \mathbf{t}_j) - p_m(\boldsymbol{\theta} \mid \mathbf{t} = \mathbf{t}_i)\} d\boldsymbol{\theta}}{\sum_{i=1}^{k-1} \sum_{j=i+1}^{k} \int (\alpha_p + \beta_p f(\boldsymbol{\theta}))\{p(\boldsymbol{\theta} \mid \mathbf{t} = \mathbf{t}_i) - p(\boldsymbol{\theta} \mid \mathbf{t} = \mathbf{t}_j)\} d\boldsymbol{\theta}}
\end{aligned}
\tag{2.2-9}
$$

Because

$$\sum_{i=1}^{k-1}\sum_{j=i+1}^{k}\int \alpha_p\{p(\boldsymbol{\theta}|\mathbf{t}=\mathbf{t}_i)-p(\boldsymbol{\theta}|\mathbf{t}=\mathbf{t}_j)+p_m(\boldsymbol{\theta}|\mathbf{t}=\mathbf{t}_j)-p_m(\boldsymbol{\theta}|\mathbf{t}=\mathbf{t}_i)\}d\boldsymbol{\theta}$$
$$=\alpha_p\sum_{i=1}^{k-1}\sum_{j=i+1}^{k}\{\int p(\boldsymbol{\theta}|\mathbf{t}=\mathbf{t}_i)d\boldsymbol{\theta}-\int p(\boldsymbol{\theta}|\mathbf{t}=\mathbf{t}_j)d\boldsymbol{\theta}+$$
$$+\int p_m(\boldsymbol{\theta}|\mathbf{t}=\mathbf{t}_j)d\boldsymbol{\theta}-\int p_m(\boldsymbol{\theta}|\mathbf{t}=\mathbf{t}_i)d\boldsymbol{\theta}\} \qquad (2.2\text{-}10)$$
$$=\alpha_p\sum_{i=1}^{k-1}\sum_{j=i+1}^{k}(1-1+1-1)$$
$$=0$$

Apply (2.2-10) to (2.2-9), we have

$$RPB=100\frac{\beta_p\sum_{i=1}^{k-1}\sum_{j=i+1}^{k}\{E(f(\boldsymbol{\theta})|\mathbf{t}=\mathbf{t}_i)-E(f(\boldsymbol{\theta})|\mathbf{t}=\mathbf{t}_j)+E_m(f(\boldsymbol{\theta})|\mathbf{t}=\mathbf{t}_j)-E_m(f(\boldsymbol{\theta})|\mathbf{t}=\mathbf{t}_i)\}}{\beta_p\sum_{i=1}^{k-1}\sum_{j=i+1}^{k}\{E(f(\boldsymbol{\theta})|\mathbf{t}=\mathbf{t}_i)-E(f(\boldsymbol{\theta})|\mathbf{t}=\mathbf{t}_j)\}} \qquad (2.2\text{-}11)$$
$$=100\frac{\sum_{i=1}^{k-1}\sum_{j=i+1}^{k}\{E(f(\boldsymbol{\theta})|\mathbf{t}=\mathbf{t}_i)-E(f(\boldsymbol{\theta})|\mathbf{t}=\mathbf{t}_j)+E_m(f(\boldsymbol{\theta})|\mathbf{t}=\mathbf{t}_j)-E_m(f(\boldsymbol{\theta})|\mathbf{t}=\mathbf{t}_i)\}}{\sum_{i=1}^{k-1}\sum_{j=i+1}^{k}\{E(f(\boldsymbol{\theta})|\mathbf{t}=\mathbf{t}_i)-E(f(\boldsymbol{\theta})|\mathbf{t}=\mathbf{t}_j)\}}$$

Therefore, the RPB is independent of p, and matching on θ alone equals the reduction in percent.

Corollary 7.2: Let $q=q(\mathrm{X})$ be some function of X. If $E(\mathrm{X}|\boldsymbol{\theta},q)=a_q+\beta_q h_q(\boldsymbol{\theta})$ for some vectors a_q and β_q and some scalar-valued functions $h_q(.)$, then total percent bias reduction at each value of q for k treatment groups can be measured by matching on θ alone at each value of q, that is

$$\frac{\sum_{i=1}^{k-1}\sum_{j=i+1}^{k} E_m(\mathrm{X}\mid \mathbf{t}=\mathbf{t}_i,q)-E_m(\mathrm{X}\mid \mathbf{t}=\mathbf{t}_j,q)}{\sum_{i=1}^{k-1}\sum_{j=i+1}^{k} E(\mathrm{X}\mid \mathbf{t}=\mathbf{t}_i,q)-E(\mathrm{X}\mid \mathbf{t}=\mathbf{t}_j,q)}=\lambda_q \qquad (2.2\text{-}1)$$

for some scalar λ_q.

Proof:

$$\frac{\sum_{i=1}^{k-1}\sum_{j=i+1}^{k} E_m(\mathrm{X}\mid \mathbf{t}=\mathbf{t}_i,q)-E_m(\mathrm{X}\mid \mathbf{t}=\mathbf{t}_j,q)}{\sum_{i=1}^{k-1}\sum_{j=i+1}^{k} E(\mathrm{X}\mid \mathbf{t}=\mathbf{t}_i,q)-E(\mathrm{X}\mid \mathbf{t}=\mathbf{t}_j,q)}$$

$$=\frac{\sum_{i=1}^{k-1}\sum_{j=i+1}^{k}\int E_m(\mathrm{X}\mid \mathbf{t}=\mathbf{t}_i,q,\boldsymbol{\theta})p_m(\boldsymbol{\theta}\mid \mathbf{t}=\mathbf{t}_i,q)d\boldsymbol{\theta}-\int E_m(\mathrm{X}\mid \mathbf{t}=\mathbf{t}_j,q,\boldsymbol{\theta})p_m(\boldsymbol{\theta}\mid \mathbf{t}=\mathbf{t}_j,q)d\boldsymbol{\theta}}{\sum_{i=1}^{k-1}\sum_{j=i+1}^{k}\int E(\mathrm{X}\mid \mathbf{t}=\mathbf{t}_i,q,\boldsymbol{\theta})p(\boldsymbol{\theta}\mid \mathbf{t}=\mathbf{t}_i,q)d\boldsymbol{\theta}-\int E(\mathrm{X}\mid \mathbf{t}=\mathbf{t}_j,q,\boldsymbol{\theta})p(\boldsymbol{\theta}\mid \mathbf{t}=\mathbf{t}_j,q)d\boldsymbol{\theta}}$$

$$=\frac{\sum_{i=1}^{k-1}\sum_{j=i+1}^{k}\int E_m(\mathrm{X}\mid q,\boldsymbol{\theta})p_m(\boldsymbol{\theta}\mid \mathbf{t}=\mathbf{t}_i,q)d\boldsymbol{\theta}-\int E_m(\mathrm{X}\mid q,\boldsymbol{\theta})p_m(\boldsymbol{\theta}\mid \mathbf{t}=\mathbf{t}_j,q)d\boldsymbol{\theta}}{\sum_{i=1}^{k-1}\sum_{j=i+1}^{k}\int E(\mathrm{X}\mid q,\boldsymbol{\theta})p(\boldsymbol{\theta}\mid \mathbf{t}=\mathbf{t}_i,q)d\boldsymbol{\theta}-\int E(\mathrm{X}\mid q,\boldsymbol{\theta})p(\boldsymbol{\theta}\mid \mathbf{t}=\mathbf{t}_j,q)d\boldsymbol{\theta}}$$

$$=\frac{\sum_{i=1}^{k-1}\sum_{j=i+1}^{k}\int (a_q+\beta_q h_q(\boldsymbol{\theta}))\{(p_m(\boldsymbol{\theta}\mid \mathbf{t}=\mathbf{t}_i,q)-p_m(\boldsymbol{\theta}\mid \mathbf{t}=\mathbf{t}_j,q)\}d\boldsymbol{\theta}}{\sum_{i=1}^{k-1}\sum_{j=i+1}^{k}\int (a_q+\beta_q h_q(\boldsymbol{\theta}))\{(p_m(\boldsymbol{\theta}\mid \mathbf{t}=\mathbf{t}_i,q)-p_m(\boldsymbol{\theta}\mid \mathbf{t}=\mathbf{t}_j,q)\}d\boldsymbol{\theta}}$$

$$=\frac{\beta_q\sum_{i=1}^{k-1}\sum_{j=i+1}^{k}\int h_q(\boldsymbol{\theta})\{(p_m(\boldsymbol{\theta}\mid \mathbf{t}=\mathbf{t}_i,q)-p_m(\boldsymbol{\theta}\mid \mathbf{t}=\mathbf{t}_j,q)\}d\boldsymbol{\theta}+a_q\sum_{i=1}^{k-1}\sum_{j=i+1}^{k}\int\{(p_m(\boldsymbol{\theta}\mid \mathbf{t}=\mathbf{t}_i,q)-p_m(\boldsymbol{\theta}\mid \mathbf{t}=\mathbf{t}_j,q)\}d\boldsymbol{\theta}}{\beta_q\sum_{i=1}^{k-1}\sum_{j=i+1}^{k}\int h_q(\boldsymbol{\theta})\{(p(\boldsymbol{\theta}\mid \mathbf{t}=\mathbf{t}_i,q)-p(\boldsymbol{\theta}\mid \mathbf{t}=\mathbf{t}_j,q)\}d\boldsymbol{\theta}+a_q\sum_{i=1}^{k-1}\sum_{j=i+1}^{k}\int\{(p(\boldsymbol{\theta}\mid \mathbf{t}=\mathbf{t}_i,q)-p(\boldsymbol{\theta}\mid \mathbf{t}=\mathbf{t}_j,q)\}d\boldsymbol{\theta}}$$

$$=\frac{\beta_q\sum_{i=1}^{k-1}\sum_{j=i+1}^{k}\int h_q(\boldsymbol{\theta})\{(p_m(\boldsymbol{\theta}\mid \mathbf{t}=\mathbf{t}_i,q)-p_m(\boldsymbol{\theta}\mid \mathbf{t}=\mathbf{t}_j,q)\}d\boldsymbol{\theta}+a_q\sum_{i=1}^{k-1}\sum_{j=i+1}^{k}(1-1)}{\beta_q\sum_{i=1}^{k-1}\sum_{j=i+1}^{k}\int h_q(\boldsymbol{\theta})\{(p(\boldsymbol{\theta}\mid \mathbf{t}=\mathbf{t}_i,q)-p(\boldsymbol{\theta}\mid \mathbf{t}=\mathbf{t}_j,q)\}d\boldsymbol{\theta}+a_q\sum_{i=1}^{k-1}\sum_{j=i+1}^{k}(1-1)}$$

$$=\frac{\sum_{i=1}^{k-1}\sum_{j=i+1}^{k} E_m(h_q(\boldsymbol{\theta})\mid \mathbf{t}=\mathbf{t}_i,q)-E_m(h_q(\boldsymbol{\theta})\mid \mathbf{t}=\mathbf{t}_j,q)}{\sum_{i=1}^{k-1}\sum_{j=i+1}^{k} E(h_q(\boldsymbol{\theta})\mid \mathbf{t}=\mathbf{t}_i,q)-E(h_q(\boldsymbol{\theta})\mid \mathbf{t}=\mathbf{t}_j,q)}$$

$$=\lambda_q$$

Corollary 7.2 implies if X define subpopulations in response to every treatment group such that some function $q = q(\mathrm{X})$ is constant within each subpopulation, then matching on θ within each subpopulation is equal percent bias reduction.

2.3 ACE given General Treatment Regime

Theorem 8 (Imai and van Dyk, 2004)

Suppose the unconfoundedness assumption holds, $\theta = \theta_{\Psi}(\mathbf{X})$, is a vector of parameters that uniquely indexes a P-Function. The ACE at θ in response between $\mathbf{t} = \mathbf{t}_a$ and $\mathbf{t} = \mathbf{t}_b$, denoted by $\tau_{a,b}$ is

$$\tau_{a,b} = \mathrm{E}\{\mathrm{Y}(\mathbf{t}_a) - Y(\mathbf{t}_b) \mid \theta\} \tag{2.3-1}$$

Proof:

$$\begin{aligned} \tau_{a,b} &= \mathrm{E}\{\mathrm{Y}(\mathbf{t}_a) \mid \mathbf{t}_a, \theta\} - \mathrm{E}\{\mathrm{Y}(\mathbf{t}_b) \mid \mathbf{t}_b, \theta\} \\ &= \mathrm{E}\{\mathrm{Y}(\mathbf{t}_a) \mid \theta\} - \mathrm{E}\{\mathrm{Y}(\mathbf{t}_b) \mid \theta\} \\ &= \mathrm{E}\{\mathrm{Y}(\mathbf{t}_a) - \mathrm{Y}(\mathbf{t}_b) \mid \theta\} \end{aligned} ,$$

where the second equality follows from Theorem 6.

In practice, when there are multiple treatment variables, we may not be interested in estimating the ACE between two treatment groups. Instead, we are more interested in estimands of average causal main effect (ACME) and average causal interaction effect (ACIE) between all groups.

2.4 Causal Inference Given Two Binary Treatments

In this section, I extend Rubin's causal model using the factorial design notations (Fisher, 1971) and P-Function theorems to estimate ACME and ACIE given two binary treatment variables. Suppose there is a vector of two independent treatments denoted as $\mathbf{T} = [T_1, T_2]$. Each treatment has two levels numbered as 0 and 1, the i the individual, in principle, has four response, which are $Y_i(0,0), Y_i(0,1), Y_i(1,0)$ and $Y_i(1,1)$. Define

$$
\begin{aligned}
Y(1,.) &= Y(1,1) + Y(1,0) \\
Y(0,.) &= Y(0,1) + Y(0,0) \\
Y(.,1) &= Y(0,1) + Y(1,1) \\
Y(.,0) &= Y(0,0) + Y(1,0)
\end{aligned}
$$

Where $Y_i(1,.) - Y_i(0,.)$ is the causal main effect (CME) of T_1 on individual i. Similarly, $Y(.,1) - Y(.,0)$ is the main causal effect of T_2 on individual i. The causal interaction effect (CIE), defined as the difference in the effect of T_1 across the levels of T_2: $\{Y(1,1) - Y(1,0)\} - \{Y(0,1) - Y(0,0)\}$. Because an individual could only receive one of four potential treatment levels combination, we cannot directly compute CME and CIE. Instead, we could estimate the ACME and ACIE. In a 2x2 factorial, experimental deign, the ACME and ACIE are defined as

$$
\begin{aligned}
\tau_{T_1} &= \mathrm{E}[Y(1,.) - Y(0,.)] = \mathrm{E}[Y(1,1)] + \mathrm{E}[Y(1,0)] - \mathrm{E}[Y(0,0)] - \mathrm{E}[Y(0,1)] \\
\tau_{T_2} &= \mathrm{E}[Y(.,1) - Y(.,0)] = \mathrm{E}[Y(1,1)] + \mathrm{E}[Y(0,1)] - \mathrm{E}[Y(0,0)] - \mathrm{E}[Y(1,0)]. \\
\tau_{T_1 \times T_2} &= \mathrm{E}[Y(1,1) - Y(1,0) - Y(0,1) + Y(0,0)]
\end{aligned}
\qquad \text{(2.4-1)}
$$

Quantities from (2.4.1) are straightforward to estimate. Because units from these groups are randomly drawn from the same population by the construction of a factorial design, and treatment assignment is independent of all baseline covariates. Nevertheless, in

observational data, lacking of randomization may lead to the systematically difference between treatment groups on pretreatment variables. Therefore, directly applying equations (2.4.1) may be misleading. Next, I derive formulas for the estimands of ACME and ACIE using P-Function theorem given STUVA and the unconfoundedness assumptions in non-experimental data.

Theorem 9: ACME and ACIE Given Two Binary Treatments

Suppose there are two independent binary treatments, $\mathbf{T} = [T_1, T_2]$. Let $\mathbf{t} = \{(0,0),(0,1),(1,0),(1,1)\}$ be a set of treatment values, and $\boldsymbol{\theta} = [\theta_1, \theta_2]$ be a vector of two PS. Under the unconfoundedness and UPPF assumptions, the ACIE between T_1, T_2 at $\boldsymbol{\theta}$, denoted by $\tau_{T_1 \times T_2}$ is given by:

$$\tau_{T_1 \times T_2} = E[Y(1,1) - Y(1,0) - Y(0,1) + Y(0,0) \mid \boldsymbol{\theta}] \tag{2.4-2}$$

The ACME of T_1 and T_2 at $\boldsymbol{\theta}$, denoted by τ_{T_1} and τ_{T_2} are

$$\begin{aligned} \tau_{T_1} &= E[Y(1,1) + Y(1,0) - Y(0,1) - Y(0,0) \mid \boldsymbol{\theta}] \\ \tau_{T_2} &= E[Y(0,1) + Y(1,1) - Y(1,0) - Y(0,0) \mid \boldsymbol{\theta}] \end{aligned} \tag{2.4-3}$$

Proof:

$$\begin{aligned} &\tau_{T_1 \times T_2} \\ &= E[Y(1,1) \mid \mathbf{t} = (1,1), \boldsymbol{\theta}] - E[Y(1,0) \mid \mathbf{t} = (1,0), \boldsymbol{\theta}] - E[Y(0,1) \mid \mathbf{t} = (0,1), \boldsymbol{\theta}] + E[Y(0,0) \mid \mathbf{t} = (0,0), \boldsymbol{\theta}] \\ &= E[Y(1,1) \mid \boldsymbol{\theta}] - E[Y(1,0) \mid \boldsymbol{\theta}] - E[Y(0,1) \mid \boldsymbol{\theta}] + E[Y(0,0) \mid \boldsymbol{\theta}] \\ &= E[Y(1,1) - Y(1,0) - Y(0,1) + Y(0,0) \mid \boldsymbol{\theta}] \end{aligned}$$

,

where the second equality follows from Theorem 6 (2.4-2).

Similarly, we have

$$
\begin{aligned}
&\tau_{T_1} = E[Y(1,1) \mid \mathbf{t} = (1,1), \boldsymbol{\theta}] + E[Y(1,0) \mid \mathbf{t} = (1,1), \boldsymbol{\theta}] \\
&-E[Y(0,1) \mid \mathbf{t} = (0,1), \boldsymbol{\theta}] - E[Y(0,0) \mid \mathbf{t} = (0,0), \boldsymbol{\theta}] \\
&= E[Y(1,1) \mid \boldsymbol{\theta}] + E[Y(1,0) \mid \boldsymbol{\theta}] \\
&-E[Y(0,1) \mid \boldsymbol{\theta}] - E[Y(0,0) \mid \boldsymbol{\theta}] \\
&= E[Y(1,1) + Y(1,0) - Y(0,1) - Y(0,0) \mid \boldsymbol{\theta}]
\end{aligned}
,
$$

and

$$
\begin{aligned}
&\tau_{T_2} = E[Y(0,1) \mid \mathbf{t} = (0,1), \boldsymbol{\theta}] + E[Y(1,1) \mid \mathbf{t} = (1,1), \boldsymbol{\theta}] \\
&-E[Y(1,0) \mid \mathbf{t} = (1,0), \boldsymbol{\theta}] - E[Y(0,0) \mid \mathbf{t} = (0,0), \boldsymbol{\theta}] \\
&= E[Y(0,1) \mid \boldsymbol{\theta}] + E[Y(1,1) \mid \boldsymbol{\theta}] - E[Y(1,0) \mid \boldsymbol{\theta}] - E[Y(0,0) \mid \boldsymbol{\theta}] \\
&= E[Y(0,1) + Y(1,1) - Y(1,0) - Y(0,0) \mid \boldsymbol{\theta}]
\end{aligned}
$$

If we take the expectation of $\tau_{T_1 \times T_2}$, τ_{T_1}, and τ_{T_2} over the distribution of $\boldsymbol{\theta}$ respectively, then

$$
\begin{aligned}
&E_{\boldsymbol{\theta}}[\tau_{T_1 \times T_2}] = E_{\boldsymbol{\theta}}[E[Y(1,1) - Y(1,0) - Y(0,1) + Y(0,0) \mid \boldsymbol{\theta}]] = E[Y(1,1) - Y(1,0) - Y(0,1) + Y(0,0)] \\
&E_{\boldsymbol{\theta}}[\tau_{T_1}] = E_{\boldsymbol{\theta}}[E[Y(1,1) + Y(1,0) - Y(0,1) - Y(0,0) \mid \boldsymbol{\theta}]] = E[Y(1,1) + Y(1,0) - Y(0,1) - Y(0,0)] \\
&E_{\boldsymbol{\theta}}[\tau_{T_2}] = E_{\boldsymbol{\theta}}[E[Y(0,1) + Y(1,1) - Y(1,0) - Y(0,0) \mid \boldsymbol{\theta}]] = E[Y(0,1) + Y(1,1) - Y(1,0) - Y(0,0)]
\end{aligned}
,
$$

where $E_{\boldsymbol{\theta}}$ denotes the expectation in regards to the distribution of $\boldsymbol{\theta}$. Thus subjects with the same value of $\boldsymbol{\theta}$ from different treatment groups could be controls for each other, and matching or stratifying on $\boldsymbol{\theta}$ can produce unbiased estimates of ACME and ACIE.

Corollary 9.1: Non-bipartite matching on P-Function Given Two Binary Treatments. Suppose there is a vector of two independent binary treatments, $\mathbf{T} = [T_1, T_2]$. Given unconfoundedness of treatment assignment and a vector $\boldsymbol{\theta}$ as balancing score that characterizes P-Function is randomly sampled from the population of individuals, units from four treatment groups $\mathbf{t} = \{(0,0), (0,1), (1,0), (1,1)\}$ are then sampled with $\boldsymbol{\theta}$. The ACME of T_i at $\boldsymbol{\theta}$, $i = 1, 2$ is the expected difference between the responses at $T_i = 1$ and

the responses at $T_i = 0$, collapsing over the levels of the other treatment variable in the matched sets. The ACIE at θ, is the expected difference of responses in the effect of T_i, across the levels of T_j in the matched set, where $i, j \in \{1,2\}, i \neq j$.

2.5 The Advantage of using Estimated P-Function over the True P-Function

In practice, the actual P-Function is unknown, and we have to estimate θ using logistic regression models, probit models, linear regressions. A number of literature (e.g., Imai and van Dyk, 2004; Rosenbaum 1987; Rubin and Thomas 1992, 1996; Hill et al,1999) proved the advantage of matching using the estimated propensity score over true propensity score. Two types of errors may explain the benefit of controlling for estimated propensity score. (1) The distribution of the treatment variables and covariates may systematically associate with each other. (2) The distribution of the covariates as a function of the treatment assignment may have random differences in an observed sample. Either the true or estimated propensity score can be used to account for the systematic relationship between the treatment and the covariates, but only the estimated propensity score can be used to account for sample-specific random differences (Rubin and Thomas 1992; Hill, Rubin, and Thomas 1999). Imai and van Dyk (2004) showed in an application study that adjusting for estimated P-Function in randomized experiments could improve estimated treatment effects.

Chapter 3

Non-bipartite Nearest Neighbor Matching

3.1 Introduction

Nearest neighbor matching (Rubin, 1974) is one of the most frequently used PS matching methods to estimate ACE in observational studies. In nearest neighbor matching, a treated unit is randomly chosen at the beginning. The control unit with the closest distance in PS is selected for matching to this treated unit. Then we repeat this procedure until untreated units are matched to all treated units. This method is also called greedy matching because the nearest control unit is chosen for matching without considering whether it is closer to a subsequent treated unit. Since nearest neighbor matching is paired matching, it may discard lots of units from the control group. Thus some researchers argued it may cause reduced power. However, the power reduction is usually minimal. Cohen (1988), and Ho et al., 2007 found the precision of ACE estimates majorly due to the smaller group size. Snedecor and Cochran (1980), verified the power increased when treated, and control groups are more similar. That is because nearest neighbor matching reduced extrapolation and increase the precision comparing to using the original data. Despite the attractiveness of this method, it is not applicable when the treatment is non-binary. Motivated by these papers, I extend the nearest neighbor matching to the multidimensional and non-bipartite scheme, namely, non-bipartite nearest neighbor matching with (NNWR) and without replacement (NN). This method is by P-Function theory addressed in chapter 2. It encompasses all treatment regime except continuous treatment variables. In Section 3.2, I describe NN and NNWR in details. In Section 3.3, I review two existing solutions: stratification on P-Function (SP) and generalized inverse probability treatment weighting (GIPTW). In

Section 3.4 I compare these methods given two binary treatment scenario using Monte Carlo simulations. In Section 3.5 I summarize the highlights in this chapter.

3.2 Methods

3.2.1 A New Distance Measure

I demonstrate a new distance that can summarize the similarity structure among multiple treatment groups. In Chapter 2 I showed matching on θ that represents a P-Function could produce an unbiased estimate of ACE. The remaining problem is how to utilize θ to match between multiple treatment groups efficiently. Let's suppose there are $t_1,...t_i,...t_g$ g disjoint subsets in total for matching. The sample size for t_i is n_i, and $n = \sum_{i=1}^{g} n_i$, where n is the total sample size. Let $t_{\min}$ be the treatment group with a smallest sample size among $t_1,...t_i,...t_g$, and the sample size of $t_{\min}$ is $n_{\min}$. Let $\theta = \theta_{\Psi}(\mathbf{X}) = \left[\theta_1,...,\theta_p\right]$ be a $p \times n$ matrix of PS estimates. Note the subscript p also corresponds to the dimension of treatment variables. For instance, given two binary treatment variables, p is 2, and g is $2 \times 2 = 4$. Under the UPPF and unconfoundedness assumption, θ uniquely represents the P-Function, where the P-Function summarizes the joint association between treatment assignment and observed covariates $\mathbf{X}$. Define d_{ij} as a Euclidean distance between a subject from subset t_i and a subject from subset t_j using θ:

$$d_{ij} = \sqrt{\sum_{u=1}^{p} (\theta_{iu} - \theta_{ju})^2} \qquad (3.2\text{-}1)$$

and $i \neq j$. Because the scales of parameter estimates can be different, it is necessary to standardize $\boldsymbol{\theta}_1,...,\boldsymbol{\theta}_p$, before applying (3.1-1). For instance, assume T_1 is dichotomous, and T_2 is categorical, where $\boldsymbol{\theta}_1$ is estimated a logistic regression model, and $\boldsymbol{\theta}_2$ using multinomial probit model, and standardize each parameter estimates before computing the distance. Suppose for $t_1,...t_i,...t_g$, and the aim is to perform a $\underbrace{1:1,...,:1:1}_{g}$ matching. In other words, each time we form a matched set of g units selected from g groups such that they are as similar to each other in the distribution of $\mathbf{X}$. In traditional setting, there are only two groups, and we can simply match a treated and an untreated subject by $d_{ij} = \left|\boldsymbol{\theta}_i - \boldsymbol{\theta}_j\right|$. However, when the size of treatment groups is greater than two, there are $C_g^2 = \frac{g(g-1)}{2}$ of possible pairwise distance. Hence using a single d_{ij} is insufficient to characterize the overall similarity when $g > 2$. To account for the similarity structure among g groups, we need to account for the closeness between units from any two groups. Therefore, I propose a distance measure denoted by D as:

$$D = \sum_{i \neq j}^{g(g-1)/2} d_{ij} = \sum_{i \neq j}^{g(g-1)/2} \sqrt{\sum_{u=1}^{p} (\theta_{iu} - \theta_{ju})^2} \qquad (3.2\text{-}2)$$

With D, we can develop efficient matching algorithms to correct for selection bias due to observed confounders among all treatment groups, and attain the approximation of the marginal distribution of potential outcomes. Figure 3.1 visualizes the calculation of D when conducting $1:1:1:1$ matching using two PS corresponding to two binary treatment variables.

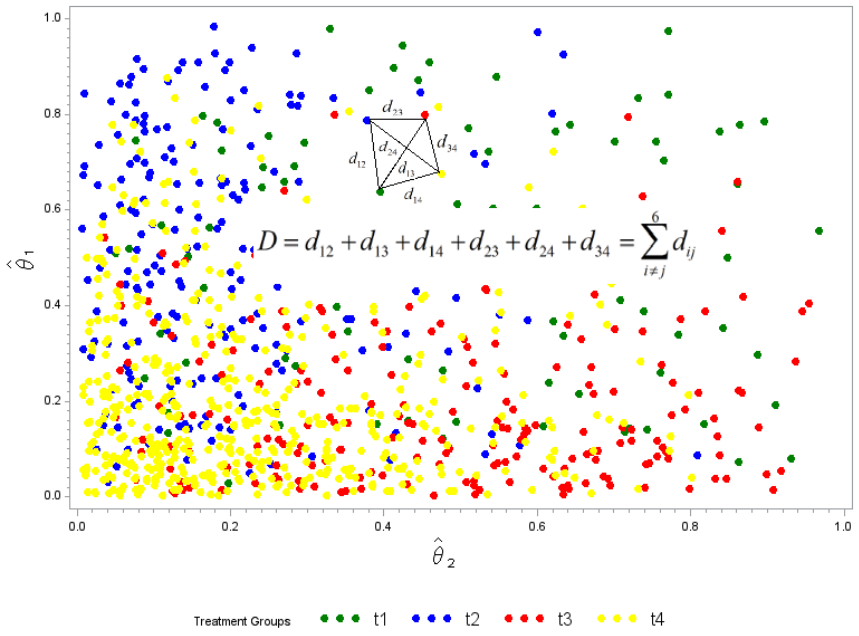

Figure 3.1 An Example of Calculating D Given Two Binary Treatment Variables

3.2.2 Non-Bipartite Nearest Neighbor Matching without Replacement (NN)

Using notations from 3.2.1, I describe the non-bipartite nearest neighbor matching without replacement strategy in following steps:

1. Transform $\boldsymbol{\theta}_1,...,\boldsymbol{\theta}_p$ to the logit scale.

2. Standardize $\boldsymbol{\theta}_1,...,\boldsymbol{\theta}_p$ if they are estimated from different types of P-Function.

3. Find the treatment group with the smallest sample size, namely $\mathbf{t}_{\min}$.

   Randomly select a subject from $\mathbf{t}_{\min}$, and search one unit from each of the other $g-1$ groups, such that the matched set minimizes $D = \sum_{i \neq j}^{g(g-1)/2} d_{ij}$.

4. Once we form a matched set, the units in it are no longer available for the remaining matching process, remove them from the reservoir of each group.

5. Repeat step 3 and step 4 until all units from $\mathbf{t}_{min}$ find their matched sets.

Step one creates a more normally distributed transformation of the propensity score. As discussed in section 1.4, matching on the logit scale of PS is better in reducing bias than matching on PS (Rosenbaum and Rubin,1985b; Rubin and Thomas,1996; Rubin, 2001). Step three that use $\mathbf{t}_{min}$ as the key treatment group is critical to achieving a better matching balance because matching units from $\mathbf{t}_{min}$ to units from other groups increase the chance of capturing a matched set with a smaller D. For example, suppose there two individuals from the group t_1 with ages 15and 20, and six individuals from t_2 with ages 50, 40, 24, 30. If we match by age and set t_2 as the reference group, the matched pairs are (15, 24) and (20, 30). Conversely, If we match from t_2 to t_1, the matched pairs are (50, 20), (40, 15). The ratio of total distance by option two and options one for two matched sets are $\frac{(50-20)}{(24-15)} \approx 3.33$, and $\frac{(40-15)}{(30-20)} = 2.5$, which are much higher. Therefore, I suggest start matching from the treatment group with the lowest sample size.

3.2.3 Non-Bipartite Nearest Neighbor Matching Method with Replacement (NNWR)

The NNWR is slightly different from NN. Matching with replacement means it allow any matched units from groups other than $\mathbf{t}_{min}$ to reenter the match candidate reservoir. By doing so, each matched set has a minimized overall distance and thus could lead to more bias reduction of ACE estimate. However, since one unit can be matched for multiple times, it could increase the variation of estimate and result in deterioration in precision of the estimate. I specify the NN approach as follows:

1. Transform $\theta_1,...,\theta_p$ to the logit scale.
2. Standardize $\theta_1,...,\theta_p$ if they are estimated from different types of P-Function.
3. Find the treatment group with the smallest sample size, namely $\mathbf{t}_{\min}$.
4. Select a subject from $\mathbf{t}_{\min}$ by the data order, and search one unit from each of the other $g-1$ groups, such that the matched set minimizes $D = \sum_{i \neq j}^{g(g-1)/2} d_{ij}$.
5. Once we form a matched set, discard the unit from $\mathbf{t}_{\min}$ that has been used in step 4.
6. Repeat step 3 and step 4 until all units the sample size from $\mathbf{t}_{\min}$ decreases to zero.
7. Once the matching process (Step 1-Step 6) is completed, assign individuals matched to $\mathbf{t}_{\min}$ group the weights equals the frequency they were selected as a match.

The major feature that NNWR different from NN is it allows to use a subject from groups rather than $\mathbf{t}_{\min}$ for matching multiple times. Therefore, it is necessary to adjust the weight for matching with replacement (Staurt, 2010; Dehejia and Wahba, 2002; Hill, Reiter, and Zanutto, 2004).

3.2.4 Data Structure After Matching

Once we obtain the matched data set using NN or NNWR with appropriate weight adjustment, the data should be structured as if they are sampled through randomization (Stuart, 2010). Austin (2007) argued whether there is a need to consider the nature of matched pairs in the outcome analysis. Schafer and Kang (2008), and Stuart (2008) pointed out it is unnecessary to account for matched pairs. The reason is PS matching cannot guarantee each pair is well-matched on the full set of covariates. In contrast, the objective of PS matching is to obtain group-

to-group level covariate balance. Therefore, it is more reasonable to aggregate all matched sets into treatment groups and perform outcome stage analysis using the groups rather than using individual level matched sets.

3.2.5 Common Support for General Treatment Regime

Using common support can exclude units with extreme values of from a treatment group that is far away from other groups, and thus help improve the quality the matched set. Flores (2009) showed using observations from treatment group in locations with extremely poor overlap in the distribution of generalized propensity scores could yield more biased ACE estimate. He further showed the bias could be reduced by using treatment groups with better common support. I extend the concept of common support to general treatment scheme. Consider there are p treatment variables and g treatment groups, the common support $\mathbf{C}$ is the overlapping region in θ among these g treatment groups. Formally: $\mathbf{C} = [\max\{\min(\theta_1,...,\theta_p)\}, \min\{\max(\theta_1,...,\theta_p)\}]$, where $\max\{\min(\theta_1,...,\theta_p)\}$ is the lower bound: the maximum value of a minimum of θ among all treatment groups, $\min\{\max(\theta_1,...,\theta_p)\}$ is the upper bound: the minimum value of maximum θ among all treatment groups.

3.2.6 Outcome Stage Analysis

Regression adjustment is the most commonly used strategy for estimating ACE after obtaining matched sample. Matching strategies are not established to "contest" with regression adjustment; actually they work best when combined with each other. (Rubin, 1973b; Rubin, 1979; Robins and Rotnitzky, 1995; Heckman et al., 1997; Rubin and Thomas, 2000; Abadie and

Imbens, 2006, Stuart, 2010). Proper matching methods are able to make the ACE estimates less sensitive to outcome model specifications (Ho et al., 2007). Regression adjustment could handle small residual covariates between treatment groups (Stuart, 2010). The modeling choice depends on the nature of the outcome. If the outcome is normally distributed, we may employ linear regression model. If the outcome is count data, we may choose generalized Poisson regression model or negative binomial model. If the outcome is time to event survival data, we may apply cox proportional hazards model (Cox., Oakes, 1984).

3.3 Two Existing Approaches

3.3.1 Stratification on P-Function (Imai and van Dyk (2004))

Instead of matching on θ, Imai and van Dyk (2004) suggested use stratification on θ. The main idea is we can equally divide each dimension of θ by a moderate number of subclasses (i.e. 3, or 4). Then within each subclass we may estimate the ACE by fitting a parametric or semiparametric model controlling for covariates significantly related to both treatment assignment and outcome. The overall ACE is then calculated by the weighted averages of within subclasses treatment effect estimate using the relative proportion of observations in each stratum as the weight. A limitation of this method is the results can be sensitive to the number of subclasses. In practice, the choice of stratifications is related to the sample size of each treatment group. A large number of strata may result in an increase of variation of treatment effects, while a small number of strata may lead to increase of bias of treatment effects. Researchers may determine the best choice by comparing the bias reduction and variance over several different options. Figure 3.2 provides an example of 3x3 stratification on two PS.

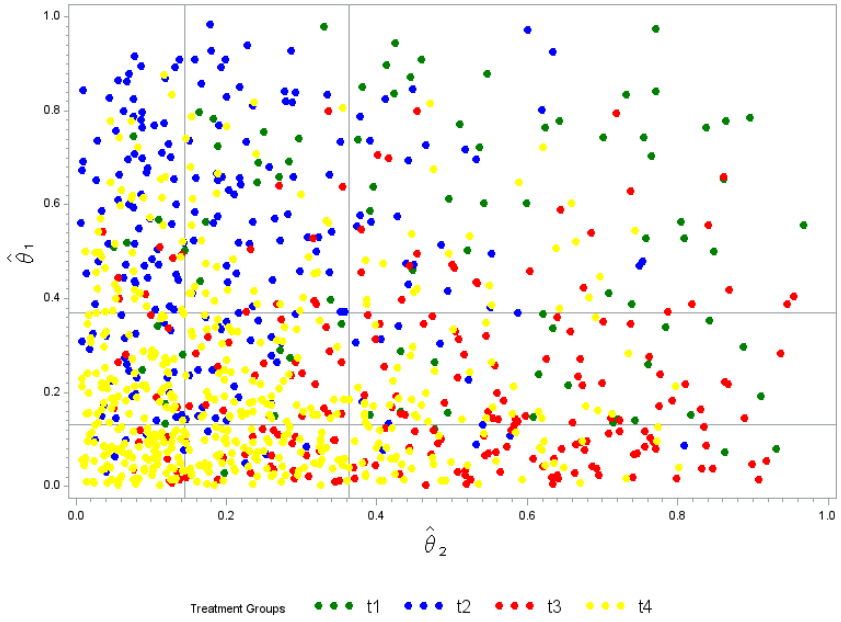

Figure 3.2 Example of 3x3 Stratification on The Distribution of Two PS

3.3.2 GIPTW (Imbens, 2000; Dong, 2015)

The GIPTW can be implemented as follows:

1. Collapse all treatment variables into one categorical variable, for example, if we have two treatment variables, one binary, and one ordinal with three levels, then we can combine them into a categorical variable with six levels.
2. Estimate GPS using multinomial logistic regression model.
3. Weight individuals who receive treatment level k by the inverse of the conditional probability of receiving the treatment they received given observed pretreatment variables.
4. Apply weighted regression model with adjustment for pretreatment covariates.

In principle, the GIPTW is used to construct a sample in which observed confounders are well adjusted and balanced among all treatment groups. The estimates generated using GIPTW

combining with regression adjustment has been shown to share a "double robustness" property, which means these estimates are consistent as long as with the outcome model or the GPS is correctly specified (Robins and Rotznizky, 1995; Scharfstein, Rotznizky, and Robins 1999)

3.4 Monte Carlo Simulations

3.4.1 Simulation Settings

To validate the performance of NN and NNWR on estimated P-Function could improve the statistical properties in the estimate of ACME and ACIE given general treatment scheme. I compare NN and NNWR with the GIPTW and Stratification on P-Function using a series of Monte Carlo simulations. I extend the one binary treatment scenario to two binary treatments scenario. The simulation setting is modified from Austin (2014). In particular, let $\boldsymbol{\theta} = \{\boldsymbol{\theta}_1, \boldsymbol{\theta}_2\}$ be a vector of the probability of treatment assignment for $\mathbf{T} = [\mathbf{T}_1, \mathbf{T}_2]$, where each parameter can be estimated from a logistic regression model. There are four treatment values: $(0,0),(0,1),(1,0),(1,1)$. We totally create 1000 datasets, in each dataset, we generate 1000 subjects along with 16 covariates ($X_1 - X_{16}$).Then we separately generate true propensity score using two logit models specified as follows.

$$\begin{aligned} &\log it(\boldsymbol{\theta}_1) = \beta_{T_{10}} + \beta_L(X_1 + X_2) + \beta_M(X_3 + X_4) + \beta_H(X_5 + X_6) + \beta_{VH}(X_7 + X_8) \\ &\log it(\boldsymbol{\theta}_2) = \beta_{T_{20}} + \beta_L(X_9 + X_{10}) + \beta_M(X_{11} + X_{12}) + \beta_H(X_{13} + X_{14}) + \beta_{VH}(X_{15} + X_{16}) \\ &X_i \sim N(0,1), \mathrm{i} = 1,...,16 \end{aligned} \qquad (3.4\text{-}1)$$

$\boldsymbol{\theta}_1(\boldsymbol{\theta}_2)$ is a vector that stores the probability of assigning $T_1(T_2)$. Each covariate follows standard normal distribution, and they are independent identically distributed. The treatment values for T_1 and T_2 are generated using two Bernoulli distributions with parameters from $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ respectively. Note the assignment of T_1 relates to $X_1,...,X_8$, while the assignment of T_2 relates

to $X_9,...,X_{16}$. Using this design, we guarantee the two treatment variables are independent of each other. I control the prevalence of $\mathbf{T}_1$ and $\mathbf{T}_2$ by changing the values of intercepts from (3.10-1) with a bisection approach (Austin, 2013). I create the continuous outcomes using the following model:

$$\begin{aligned} Y(\mathbf{t}) &= \beta_L(X_1 + X_2 + X_9 + X_{10}) + \beta_M(X_3 + X_4 + X_{11} + X_{12}) \\ &+ \beta_H(X_5 + X_6 + X_{13} + X_{14}) + \beta_{VH}(X_7 + X_8 + X_{15} + X_{16}) + \tau \mathrm{T}_1 + \tau \mathrm{T}_2 + \tau \mathrm{T}_1 \times \mathrm{T}_2 + \boldsymbol{\varepsilon} \end{aligned} \tag{3.4-2}$$

, where $\boldsymbol{\varepsilon} \sim N(0,3)$. The true ACME and ACE denoted by τ are set to 1. The values of regression coefficients are specified as: $\beta_L = \log(1.25), \beta_M = \log(1.5), \beta_H = \log(1.75), \beta_{VH} = \log(2)$, where the subscripts L, M, H and VH indicate Low, Moderate, High, and Very High. Hence there are two covariates with low effect on treatments assignments of T_1 (T_2), two covariates with moderate effect on treatments assignments of T_1 (T_2), two covariates with high effects on treatments assignments of T_1 (T_2), and two covariates with very strong effects on treatments assignments of T_1 (T_2). Four covariates have low effect, four covariates have moderate effects, four covariates have high effects, and four covariates have very high effects on outcomes. We set prevalence of receiving T_1 and T_2 as addressed in Table 3.1. The first column lists prevalence of T_1, and the first row lists prevalence of T_2. The cell with "Yes" indicates that scenario is included in the simulation. Totally there are 31 prevalence cases. For instance, if the prevalence of T_1 and T_2 are 0.2,0.3, and the total sample size is 1000, then the sample size of treatment groups receiving T_1 (T_2) is 200 (300). The order of prevalence for T_1 and T_2 has no consequence. Hence I remove certain scenarios due to redundancy (e.g. the scenario when the

prevalence of T_1 is 0.8 and T_2 is 0.2 is equivalent to the scenario when the prevalence of T_1 is 0.2 and T_2 is 0.8).

Table 3.1. Summary for Prevalence of T_1 and T_2

T_1/T_2	0.2	0.3	0.4	0.5	0.6	0.7	0.8
0.2	Yes	Yes	Yes	Yes	Yes	Yes	Yes
0.3	Yes	Yes	Yes	Yes	Yes	Yes	Yes
0.4	Yes	Yes	Yes	Yes	Yes	Yes	Yes
0.5	Yes	Yes	Yes	Yes	Yes	Yes	Yes
0.6	No	No	No	No	Yes	No	No
0.7	No	No	No	No	No	Yes	No
0.8	No	No	No	No	No	No	Yes

In each replication, I estimate the two PS $\boldsymbol{\theta} = \{\boldsymbol{\theta}_1, \boldsymbol{\theta}_2\}$ using logistic regression model as follows:

$$\begin{aligned} \log it(\boldsymbol{\theta}_1) &= \boldsymbol{\beta}_{T_{10}} + \boldsymbol{\beta}_{T_{11}} \boldsymbol{X} \\ \log it(\boldsymbol{\theta}_2) &= \boldsymbol{\beta}_{T_{20}} + \boldsymbol{\beta}_{T_{21}} \boldsymbol{X} \end{aligned} \qquad (3.4\text{-}3)$$

, where $\boldsymbol{X}$ includes all 16 covariates related to outcome. The model specifications in 3.4-3 $\boldsymbol{\theta} = \{\boldsymbol{\theta}_1, \boldsymbol{\theta}_2\}$ are different from the true models specified in (3.10-1) due to two reasons. First, these misspecifications mimic the situation when the correct model is unknown in reality. Second, the previous studies suggested covariates related to outcome should always be included in the model for PS estimation no matter if they are associated with treatment assignment (Brookhart et al. ,2006; Austin, Grootendorst, and Anderson, 2007).

After estimating $\theta = \{\theta_1, \theta_2\}$, I trim the raw data using the common support (see section 3.2.5). I use the trimmed data for all PSM. I obtain matched sample using NN and NNWR to prepare for outcome stage modeling. In the matched sample generated by NN, I approximate the ACME and ACIE addressed in (2.4-1) using linear regression controlling all 16 covariates related to the response variable. In the matched sample by NNWR, I apply the weighted linear regression. The weight of a unit is the frequency the unit is matched. When using stratification on P-Function (SP), I equally divide the distribution of θ_1 into three strata and do the same to θ_2, namely, 3x3 SP. Within each stratum, I estimate the ACME and ACIE using linear regression model controlling for X_1 through X_{16}. Then I calculate the overall ACME and ACIE by taking the weighted averages of within-subclass ACME and ACIE estimates. The weight of a stratum is the ratio of its sample size to the overall sample size. When using GIPTW, I follow the steps specified in Section 3.5. I use relative bias (RB) in estimating the ACME and ACIE (3.4-4) and mean square error (MSE) (3.4-5) of estimated ACME and ACIE to examine the performance of these methods.

$$\mathrm{RB}(\hat{\tau}) = \frac{\sum_{i=1}^{1000} (\frac{\hat{\tau}_i}{\tau} - 1)}{1000} \tag{3.4-4}$$

$$\mathrm{MSE}(\hat{\tau}) = \sum_{i=1}^{1000} (\hat{\tau}_i - \tau)^2 \tag{3.4-5}$$

$\hat{\tau}_i$ represents the ACME or ACIE estimate for the i the replication, and τ denotes the true ACME or ACIE.

3.4.2 Simulation Results

Case: Prevalence of T_1 and T_2 are equal

Figure 3.3 presents the MSE and RB of estimated ACIE and ACME using different matching algorithms across seven cases where the prevalence of T_1 and T_2 are ranged from 0.2 to 0.8 by 0.1.

In regards to the MSE of ACIE estimates, the NN algorithm tended to result in lowest MSE among four methods; The NNWR algorithm yielded much higher MSE than other three methods. GIPTW is slightly better than 3×3 SP when the prevalence of T_1 and T_2 are from 0.2 to 0.5. The discrepancies between any two approaches tended to decrease as the prevalence of T_1 and T_2 simultaneously increased from 0.2 to 0.5. All methods resulted in lowest MSE of ACIE when the prevalence of T_1 and T_2 are $0.5, 0.5$. The plot is almost symmetric to $0.5, 0.5$ (e.g. the MSE of ACIE for each method where the prevalence of T_1 and T_2 are 0.3 is closed to MSE to the same method where the prevalence of T_1 and T_2 are 0.7). The NN, GIPTW, and 3x3 SP have very similar performance when the prevalence of T_1 and T_2 are 0.5.

The MSE of ACME estimates for T_1 and T_2 are relatively similar on each method. The NN algorithm is superior to other three methods when the prevalence of T_1 and T_2 are 0.5 to 0.8. The NNWR algorithm is the worst among four methods. The GIPTW tended to result in smaller MSE of ACME estimates when the prevalence of T_1 and T_2 are from 0.2 to 0.3. The NN, GIPTW, and 3x3 SP have very similar performance when the prevalence of T_1 and T_2 are from 0.4 to 0.5. The disparities between NN and other three methods tended to increase as the prevalence of T_1 and T_2 simultaneously rose from 0.5 to 0.8.

In regards to the RB of ACIE estimates, all methods except the 3x3 SP in some settings produced RB less than 10%. Specifically, RB tended to be less and closed to be unbiased with NN algorithm when the prevalence of T_1 and T_2 are $0.2, 0.2$ and $0.7, 0.7$. The NNWR produced lower RB than NN when the prevalence of T_1 and T_2 are $0.3, 0.3$ and $0.5, 0.5$. The GIPTW produced smaller RB than other methods when the prevalence of T_1 and T_2 are $0.5, 0.5$ and $0.8, 0.8$. The 3x3 SP tended to generate lowest RB when the prevalence of T_1 and T_2 are $0.4, 0.4$ and $0.6, 0.6$. The change in RB for NN is stable (less than 3%) when ignoring the case for $0.3, 0.3$.

Regarding the RB of ACME estimates for T_1, all methods except the 3x3 SP in the cases of $0.7, 0.7$ and $0.8, 0.8$ tended to result in RB less than 5%. NNWR is better than other methods in the cases of $0.2, 0.2$; $0.4, 0.4$ and $0.5, 0.5$. The change in RB for NN, NNCW, and GIPTW are less than %3 across all cases. 3x3 SP tended to result in much greater RB in the cases of $0.7, 0.7$ and $0.8, 0.8$. The results regarding RB of ACME estimates for T_2 is similar to the reuslts for T_1. NN is tended to result in lowest RB in the cases of $0.4, 0.4$. The NNWR outperformed other methods but is closed to NN in the cases of $0.6, 0.6$ and $0.7, 0.7$. RB tended to be least with GIPTW in the cases of $0.2, 0.2$; $0.3, 0.3$ and $0.5, 0.5$. 3x3 SP is inferior to other approaches in the cases of $0.7, 0.7$ and $0.8, 0.8$. In general, when the prevalence of two treatment variables is the same, the NN algorithm is favored than other three methods regarding lowest MSE of ACIE and ACME in most of the scenarios. In addition, the RB of ACIE and ACME of NN is less than 5% in most of the cases.

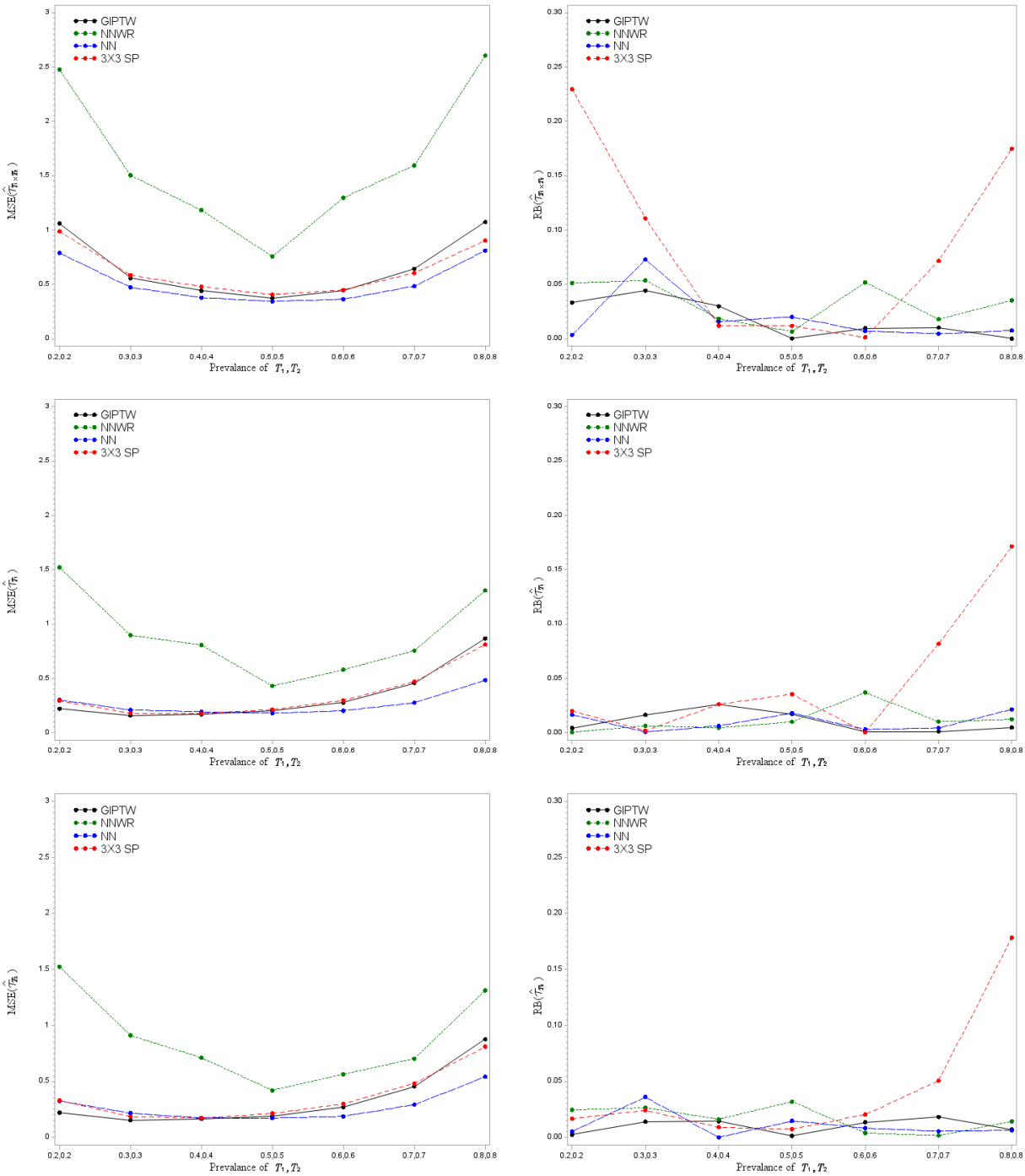

Figure 3.3 Case: Prevalence of T_1 and T_2 are equal

However, the NN with lowest MSE of ACIE does not necessarily have the lowest ACME in some settings (i.e. the cases of $0.2, 0.2$ and $0.3, 0.3$). Finally, although the RB of ACIE and ACME is the lowest for NNWR methods in some contexts, the NNWR showed the poorest performance with largest MSE as well as largest variation among four methods.

Case: Prevalence of T_1 equals 0.2 and prevalence of T_2 unequal 0.2

Figure 3.4 depicts the MSE and RB of estimated ACIE and ACME using different matching algorithms across six cases where the prevalence of T_1 is 0.2 and prevalence of T_2 is from 0.3 to 0.8 by 0.1.

In regards to MSE of ACIE estimates. The MSE for NN is about 33% less than the MSE for both GIPTW and 3x3 SP and is roughly over 70% less than the MSE for NNWR. The performance of 3x3 SP is very closed to GIPTW as the prevalence of T_2 increased from 0.3 to 0.7, but better than GIPTW when the prevalence of T_2 equals 0.8. The magnitude of MSE for NN, GIPTW, and 3x3 SP diminished as the prevalence of T_2 decreased from 0.8 to 0.2, and almost coincided as the prevalence of T_2 decreased from 0.4 to 0.2. The NNWR still resulted in the highest MSE across all six scenarios.

In regards to MSE of ACME estimates for T_1, NN intended to result in lowest MSE comparing to other methods. GIPTW and 3x3 SP exhibited slight disparities as the prevalence of T_2 increased from 0.3 to 0.7. The magnitude of MSE for NN is steady when the prevalence of T_2 is between 0.3 and 0.7. When the prevalence of T_2 is 0.8, the MSE is the highest among six scenarios on each method.

With respect to MSE of ACME estimates for T_2, the GPTW intended to yield slightly lower MSE than those of NN and 3x3 SP, but much lower MSE than the one of the NNWR. The curves in response to 3x3 SP and NN are almost overlapped one to another. The slopes for all methods other than NNWR are nearly flat when the prevalence of T_2 is between 0.3 and 0.7.

In regards to RB of ACIE estimates, all methods except 3x3 SP intended to result in RB less than 8%. The NN tended to result in lower RB compared to other methods when the prevalence of T_2 is between 0.5 to 0.8. The NNWR tended to produce smaller RB than other methods when the prevalence of T_2 is 0.4. The GIPTW is superior to other methods when the prevalence of T_2 is 0.3. The magnitude of MSE due to NN is the most stable among all approaches with bias less than 5% across all six cases.

With respect to RB of ACME estimate for T_1, all methods except 3x3 SP in two settings tended to result in RB less than 7%. The NN tended to result in lower RB compared to other methods in the cases of $0.2,0.3;0.2,0.5;0.2,0.7$ and $0.2,0.8$. The RB due to NNWR is closed to NN in the cases of 0.2,0.3; 0.2,0.6 and 0.2,0.7. The RB due to GIPTW is pretty similar to those for NN except the case of 0.2,0.7. The RB due to 3x3 SP is the greatest across all six scenarios.

In regards to RB of ACME estimates for T_2. All methods except NNWR in two settings (0.2,0.5 and 0.2,0.8) had similar RB. NN, 3x3 SP, and GIPTW tended to yield RB less than 3% across all

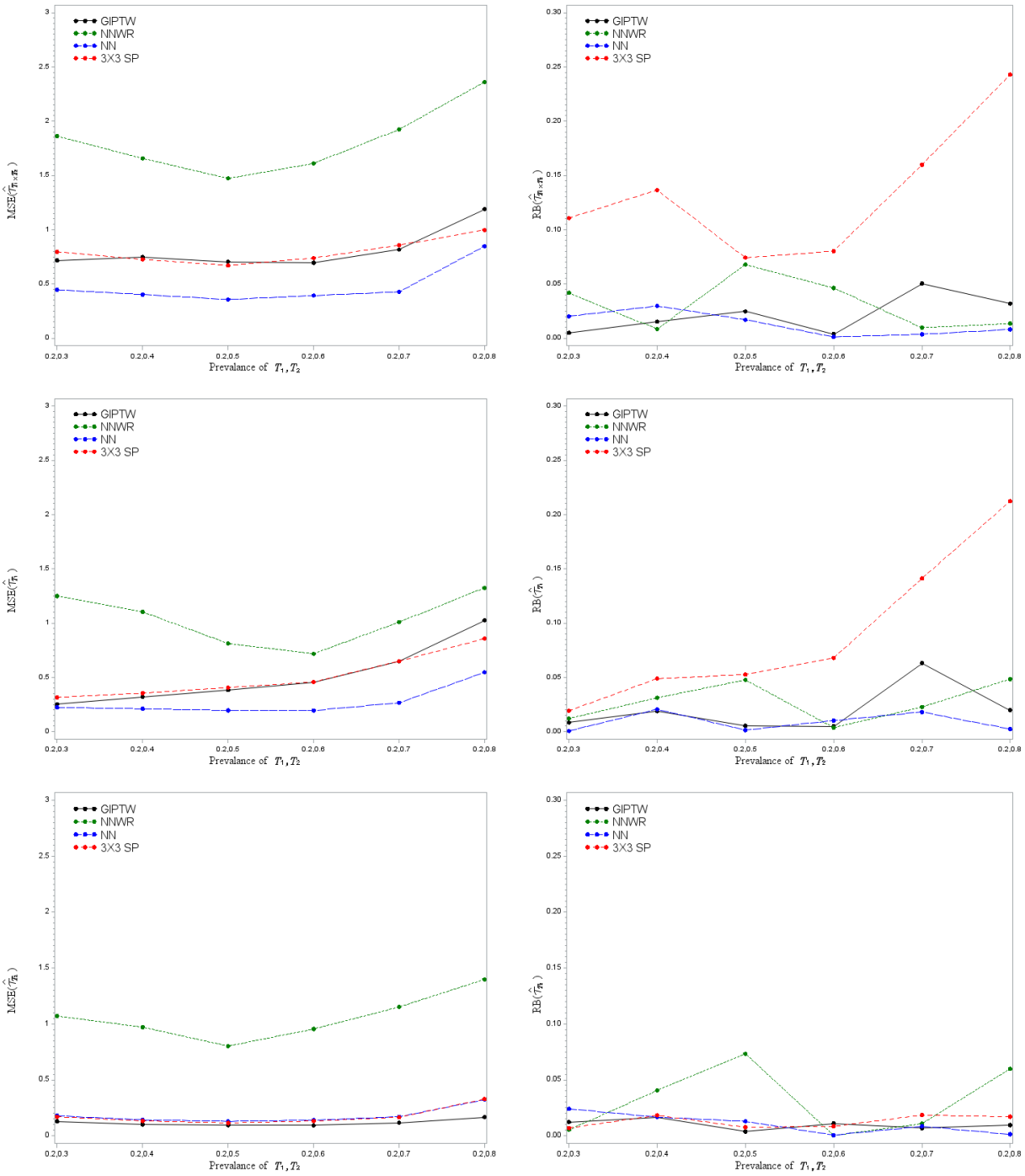

Figure 3.4 Case: Prevalence of T_1 is 0.2 and not equal to Prevalence of T_2

cases. RB due to NN steadily decreased as the prevalence of T_2 increased from 0.2 to 0.6 and stabilized from 0.6 to 0.8.

In general, when Prevalence of T_1 equals 0.2 and prevalence of T_2 is not equal to 0.2, the NN algorithm is better than other three methods regarding lowest MSE of ACIE and ACME in most of the scenarios. Also, the RB of ACIE and ACME of NN is less than 5% in most of the cases. The NNWR in inferior to other methods with largest MSE among four methods. The RB of ACME estimates due to 3x3 SP is sensitive to the prevalence settings in the sense that it yielded over 10% RB in certain cases.

Case: Prevalence of T_1 equals 0.3 and prevalence of T_2 unequal 0.3 (Figure 3.5)

The MSE and RB of ACIE and ACME plots seem pretty similar to the previous case where the prevalence of T_1 is 0.2. I only highlight the major patterns different from previous cases. In regards to the MSE of ACME estimates for T_1. The two curves of 3x3 SP and GIPTW are almost overlapped. In regards to the RB of ACIE, all methods except the 3x3 SP tended to result in RB less than 8%. The NN tended to produce smallest RB in all scenarios except when the prevalence of T_2 is 0.4. The 3x3 SP tended to yield highest RB in the cases of 0.3,0.2; 0.3,0.6; 0.3,0.7; and 0.3,0.8.

Case: Prevalence of T_1 equals 0.4 and prevalence of T_2 unequal 0.4 (Figure 3.6)

The MSE and RB of ACIE and ACME plots look very comparable to the previous case where the prevalence of T_1 is 0.2 and 0.3.

The MSE and RB of ACIE and ACME plots seem pretty similar to the previous case where the prevalence of T_1 is 0.2. I only address the key patterns different from previous cases. RB of ACIE estimates due to NNWR is higher than the RB of ACIE estimates due to other

methods. RB of ACME estimates due to NNWR is the largest when the prevalence of T_2 is between 0.4 and 0.8.

Case: Prevalence of T_1 equals 0.5 and prevalence of T_2 unequal 0.5 (Figure 3.7)

We observe similar patterns in the case when the prevalence of T_1 equals 0.4. RB of ACIE estimates due to NNWR is the lowest when the prevalence of T_2 is 0.2 and 0.6. RB of ACIE estimates due to NN is the lowest when the prevalence of T_2 is 0.4 and 0.8. RB of ACME estimates for T_1 due to 3x3 SP is the lowest when the prevalence of T_2 is 0.3, 0.4, and 0.6. RN of ACME estimates for T_2 due to NNWR is the lowest when the prevalence of T_2 is 0.7.

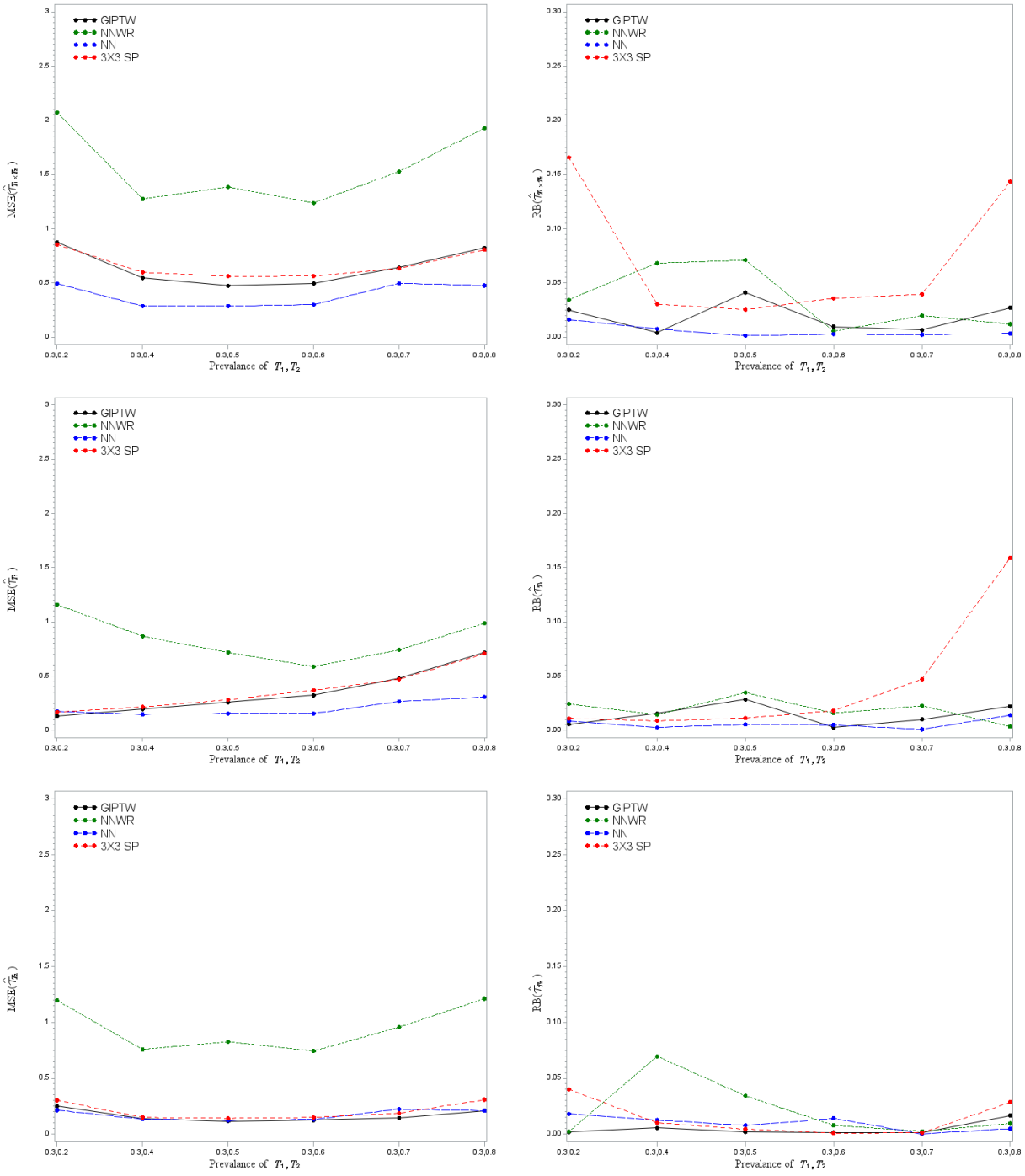

Figure 3.5 Case: Prevalence of T_1 is 0.3 and not equal to Prevalence of T_2

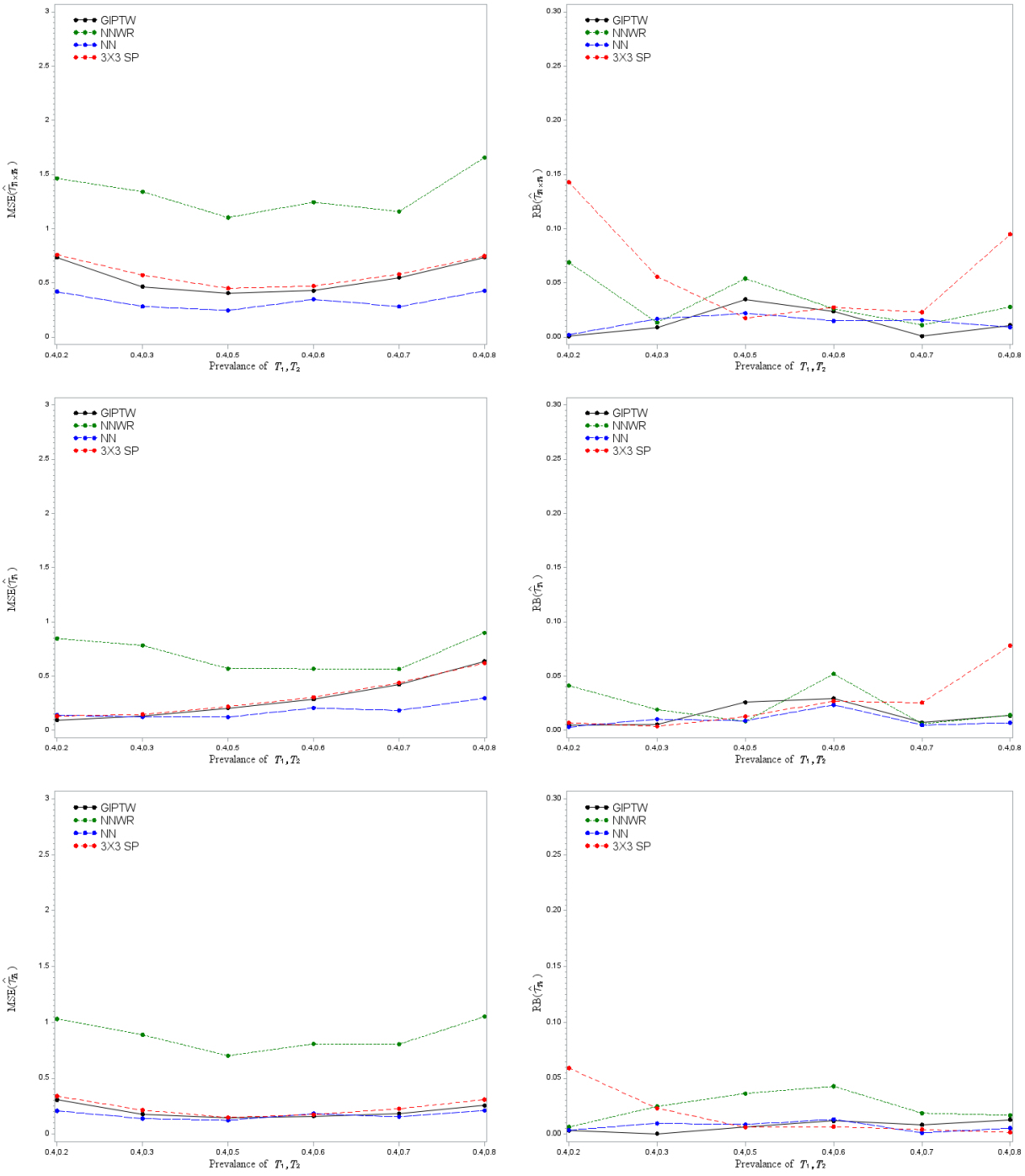

Figure 3.6 Case: Prevalence of T_1 is 0.4 and not equal to Prevalence of T_2

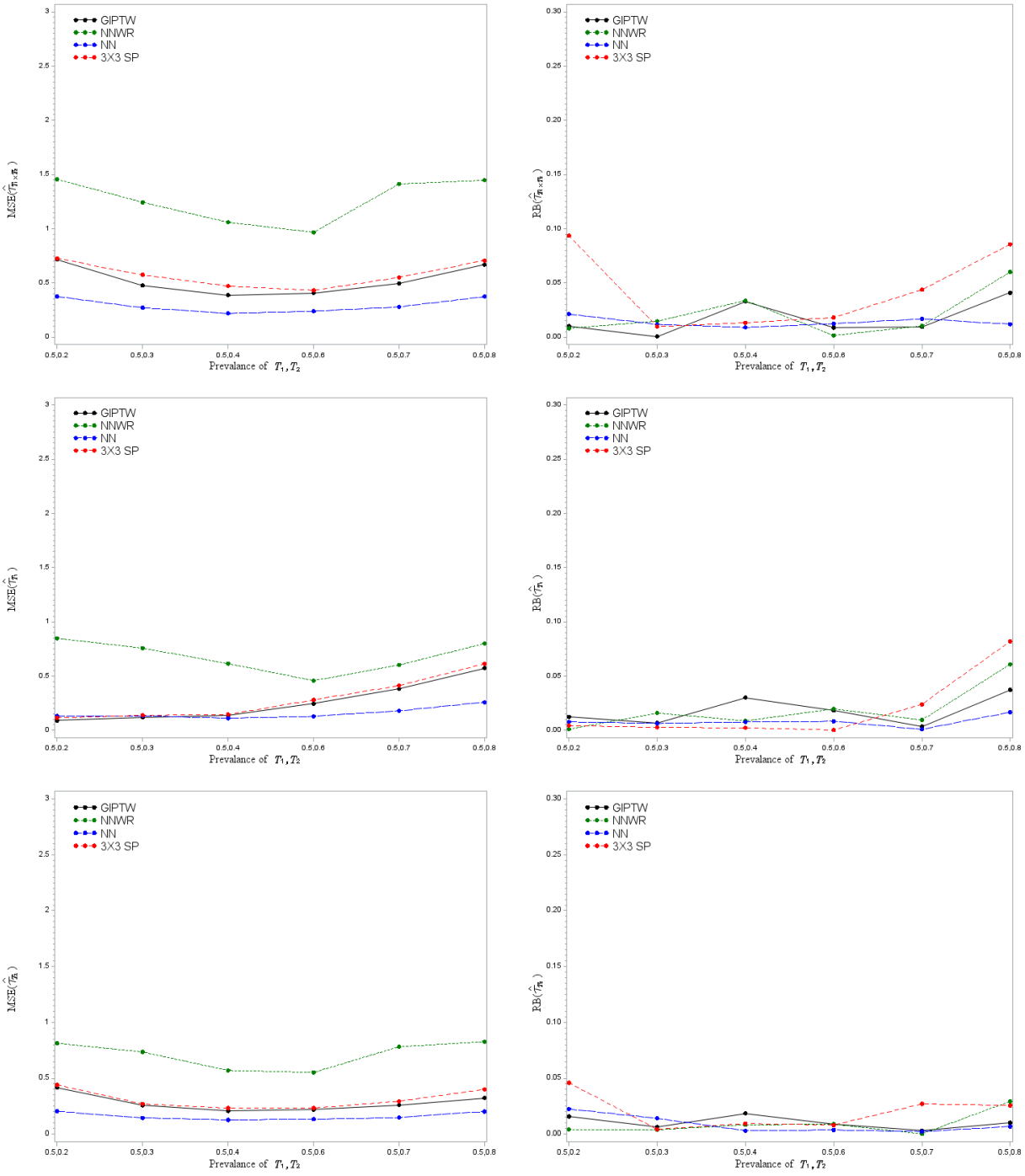

Figure 3.7 Case: Prevalence of T_1 is 0.5 and not equal to Prevalence of T_2

3.4 Discussion

Estimating ACIE on the continuous outcome is essential when a study involves multiple treatment variables. In this study, I extended the nearest neighbor matching with and without replacement on PS to non-bipartite nearest neighbor matching with (NN) and without replacement (NNWR) on P-Function. Both NN and NNWR can be applied to general treatment scheme except the continuous treatment. I used an extensive series of Monte Carlo simulations to evaluate the ability of proposed two methods and other two popular methods to estimate the ACIE and ACME given two binary treatment variables. Of the different PSM examined, the NN resulted in the lowest MSE when estimating the ACIE across all 31 scenarios. Besides, the NN led to the lowest MSE when estimating the ACME of both two treatment variables in over 74% scenarios. Another advantage of NN is the MSE, and RB of ACIE and ACME estimates are approximately constant across all cases. Unlike NN, the performance of other three methods are associated with the prevalence of treatments. For instance, the MSE of ACME on T_2 due to 3x3 SP and IPTW substantially increased when the prevalence of T_1 is between 0.2 to 0.5 and prevalence of T_2 is greater than 0.4 and not equal to the prevalence of T_1. The NNWR is inferior to other methods when estimating MSE of ACIE and ACME. The GIPTW produced lower MSE of both ACME and ACIE than 3x3 SP and NNWR in most of the cases, and had comparable performance to NN when examining the RB of ACIE and ACME estimates. The 3x3 SP had comparable MSE of ACIE and ACME estimates to the GIPTW in the majority of

cases, but substantially higher RB in certain cases. This finding implies the variance of estimates due to 3x3 SP is lower than the GIPTW.

When the primary objective is to estimate the ACIE between two treatment variables, I suggest the NN method be used. When the main purpose is to estimate the ACME of each treatment, then the choice depends on the prevalence of treatment and which ACME the researchers are more interested in. For example, when estimating the ACME of T_2 and the prevalence of T_1 is 0.2, I would suggest use GIPTW. Otherwise, the NN is the best solution.

There were a few of limitations to the simulation study. First, I did not consider comparing the sensitivity of these methods to the misspecification parametric model in estimating P-Function. I did not investigate this issue because the simulations were extensive with essentially 61 scenarios. Introducing another factor would have substantially increased a number of results. Another limitation is the simulation only include continuous outcome. In subsequent research, evaluating these methods in estimating ACIE and ACME given binary, ordinal, or survival outcome merits further investigation. A third limitation is the NNWR and NN can be further generalized to be able to apply to continuous treatment. A fourth limitation is both NNWR and NN did not consider matching one unit with a fixed number of units or a variable number of units in each of other treatment groups. In subsequent research, it is desirable to check whether these changes could further improve these methods.

Chapter 4.

The Non-Bipartite Nearest Neighbor Caliper Width Variable Matching

4.1 Introduction

Propensity score matching methods (PSM) (Rosenbaum and Rubin, 1983) are increasingly employed to reduce treatment selection bias when estimating average causal effects in non-experimental studies. Propensity score (PS) is defined as the conditional probability of being treated given observed baseline characteristics. Rubin and Rosenbaum (1983) showed given the unconfoundedness, and STUVA assumption, matching or stratifying on PS could replicate a randomization process such that the distribution of baseline covariates are comparable between treated and untreated subjects. The bipartite nearest neighbor caliper width matching was found the most commonly used PSM in the medical literature (Austin, 2009). The nearest neighbor caliper width matching was modified from the nearest neighbor matching by imposing a specified caliper width. The caliper width was defined as a weighted pooled standard deviation of PS or logit of PS from treatment and control groups. With this algorithm, a treated subject can find an untreated subject only if their distance in PS is within the caliper width. Clearly, with a tight caliper width, one treated subject might not be able to find a match if the distance to its nearest neighbor is beyond the caliper width. Therefore, tight caliper width is preferred when there is a large reservoir of controls. Otherwise, a loose caliper width can be better (Mark, 2013). The nearest neighbor matching is essentially a particular case of nearest caliper width matching.

Because as the caliper width increase to infinity, the restriction on caliper width no longer exists. The performance of nearest neighbor caliper width matching relies on the choice of caliper width. Cochran and Rubin (1973) showed matching on a normally distributed confounding variable with 0.2, 0.4, 0.6, 0.8, and 1 of the standard deviation of the confounder distribution removed approximately 99%, 95%, 89%, 82%, and 74% of the bias due to the confounder. Rubin and Rosenbaum (1985) examined the performance of this approach in estimating treatment effect when matching on PS. They found using 0.2 and 0.6 of the standard deviation of pooled logit PS eliminated nearly 99% and 89% of the bias due to observed pretreatment variables. Austin (2009) did a comprehensive literature review on nearest neighbor caliper width matching on PS in medical research. He found Ayanian (2002) used 0.6 of a standard deviation of the logit of PS. Some literature used 0.005 of standard deviation of PS (Iwashyna and Lamont, 2002; Cole et al., 2002; Christakis and Iwashyna, 2003), 0.01 (Weiss et al., 2002; Ferguson, Coombs and Peterson, 2002; Seeger et al., 2003; Magee et al., 2003; Hall, Sumners and Obenchain, 2003). Austin (2011) conducted a simulation study to determine the optimal caliper width when using nearest neighbor caliper width matching under various settings. He recommended using 0.2 of the standard deviation of logit propensity score that would minimize mean square error in a majority of scenarios. Austin (2014) used this optimal caliper width matching algorithm and compared it with other 11 matching algorithms including nearest neighbor matching and optimal matching. He found the nearest neighbor optimal caliper (0.2) width matching had the best performance regarding lowest mean square error.

Despite the popularity of nearest neighbor caliper width matching, the major constraint is it cannot be applied to non-binary treatment variables. When there is one binary treatment variable, the caliper width is formulated as:

$$C = w\sqrt{\frac{(s_1^2 + s_0^2)}{2}} \tag{4.1-1}$$

where w is a user-defined weight, S_1^2 is the sample variance of the propensity score in the treatment group, and S_0^2 is the sample variance of the propensity score in the control group. However, with the multiple treatment cases, this formula is ineligible. For example, if there are two categorical treatment variables, we may estimate the P-Function using bivariate multinomial logistic regression, and characterize the P-Function of each subject by a vector of two generalized propensity scores. If we directly use (4.1-1), both s_1^2 and s_0^2 becomes a 2x2 matrix, and the caliper width C also becomes a 2x2 matrix rather than a scalar. Hence it cannot be used as a caliper width for a distance in scalar form.

Another direction to improve the nearest neighbor matching was matching each treated units with a variable number of controls (Ming and Rosenbaum, 2000). Nearest neighbor matching allows one treated unit to match with a fixed number of controls. Conversely, Ming and Rosenbaum (2000) found in certain scenarios matching with a variable number of controls could substantially eliminate more bias than matching with a fixed number of controls. Moreover, previous literature only paid attention to the nearest neighbor caliper width without replacement but ignore the potential of nearest neighbor caliper width matching with replacement. Motivated by these developments and restrictions, we aim to develop new methodologies that incorporate the advantages of these two methods and extend them to artificial treatment scheme. I propose two methodologies: non-bipartite nearest neighbor caliper width variable matching with replacement (NNCVWR) and without replacement (NNCV) on P-Function. In section 4.2 I introduced NNCVWR and NNCV methods. In section 4.3, I conduct a series of Monte Carlo simulation to find out the optimal caliper widths on a variety of treatment

assignment settings. To examine the performance of these two methods, I compare NNCVWR and NNCV with optimal caliper width to the NN, GIPTW and 3x3 SP illustrated in chapter 3 using the simulated data. In section 4.4, I present the simulation results and provide guidance in choosing optimal caliper width in practice.

4.2 Methods

4.2.1 A New Caliper Width

I use the notations in section 3.2.1. Assume there are totally n individuals in a sample, g disjoint treatment groups denoted by $t_1,...t_i,...t_g$. n_i is a sample size of the group t_i, such that $n = \sum_{i=1}^{g} n_i$. Let $t_{\min}$ denote the group with smallest sample size among $t_1,...t_i,...t_g$, and $n_{\min}$ is the sample size of $t_{\min}$. Let $\boldsymbol{\theta} = \boldsymbol{\theta}_{\Psi}(\mathbf{X}) = \left[\boldsymbol{\theta}_1,...,\boldsymbol{\theta}_p\right]$ be a $p \times n$ matrix that uniquely indexes the P-Function. Recall in section 3.2 we defined the "closeness" between a unit from the group t_i and a unit from the group t_j as $d_{ij} = \sqrt{\sum_{u=1}^{p} (\theta_{iu} - \theta_{ju})^2}$ (see 3.2-1). Let $\mathbf{d}_{ij}$ be a set including all possible d_{ij}, $d_{ij} \in \mathbf{d}_{ij}$, and s_{ij} be the standard deviation of $\mathbf{d}_{ij}$. The new caliper width corresponding to t_i and t_j is

$$C_{ij} = w \cdot s_{ij} \tag{4.2-1}$$

where w is a customized weight. Note here w is a constant and does not change when the treatment groups vary. Using this new caliper width, I introduce the NNCV and NNCVWR in the following two sections.

4.2.2 Non-Bipartite Nearest Neighbor Caliper Width Variable Matching without Replacement (NNCV)

Using notations from 4.2.1, I elaborate the NNCV in 10 steps:

1. Transform $\theta_1,...,\theta_p$ to the logit scale.
2. Standardize $\theta_1,...,\theta_p$ if they are estimated from different types of P-Function.
3. Find the group $\mathbf{t}_{\min}$ with the lowest sample size among $t_1,...t_i,...t_g$. Randomly select $t'_{\min} \subseteq t_{\min}$. $t'_{\min}$ includes a single unit u from $t_{\min}$, $u \in t_{\min}, t'_{\min} \subseteq t_{\min}$.
4. Define $\{t'_1,...,t'_i,...,t'_{g-1}\} \subseteq \{t_1,...,t_i,...,t_{g-1},t_i \neq t_{\min}\}$, where t'_i is a possible subset of t_i, and $\{t'_1,...,t'_i,...,t'_{g-1}\}$ is a subsample of $\{t_1,...,t_i,...,t_{g-1}\}$. For the unit u from $t'_{\min}$, find a unique candidate reservoir: $\{t'_1,...,t'_i,...,t'_{g-1}\}$ corresponding to u, such that the distance between any unit from t'_i and any unit from t'_j, $\{t'_i,t'_j,i \neq j\} \in \{t'_{\min},t'_1,...,t'_i,...,t'_{g-1}\}$ meets the criteria:

$$d_{ij} \leq C_{ij} = w \cdot s_{ij} \tag{4.2-2}$$

   If no candidates meet (4.2-2), stop and return to step 3.
5. Match the unit in $t'_{\min}$ with one unit from each of other $g-1$ groups $t'_1,...,t'_i,...,t'_{g-1}$ such that their overall distance minimize

$$D = \sum_{\{t'_i,t'_j,i\neq j\}\in\{t'_{\min},t'_1,...,t'_i,...,t'_{g-1}\}}^{g(g-1)/2} d_{ij} \tag{4.2-3}$$

   Note (4.2-3) is a special case of (3.2-2).
6. Keep the unit in $t'_{\min}$. Remove the nearest neighbors selected from $t'_1,...,t'_i,...,t'_{g-1}$ in step 5.

7. Repeat Step 5 and step 6 up to m times, such that the unit in $t'_{\min}$ is matched with up to m units from each of $t'_1,...,t'_i,...,t'_{g-1}$. Save the matched sets and remove $t'_{\min}$ from $t_{\min}$.
8. Let m'_u be the actual number of units selected from t'_i and matched to unit u, $t'_i \in \{t'_1,...,t'_i,...,t'_{g-1}\}$. Set the weight of all unit matched to unit u as $\frac{1}{m'_u}$.
9. Repeat step 3 to 8 until all units from $t_{\min}$ has been selected.
10. Aggregate all matched sets into treatment groups for outcome stage analysis.

4.2.3 Non-Bipartite Nearest Neighbor Caliper Width Variable Matching with Replacement (NNCVWR)

Similar to NNCV, the NNCVWR can be implemented as follows:

1. Transform $\theta_1,...,\theta_p$ to the logit scale.
2. Standardize $\theta_1,...,\theta_p$ if they are estimated from different sorts of P-Function.
3. Find the group $\mathbf{t}_{\min}$ with the smallest sample size among $t_1,...t_i,...t_g$. Randomly choose $t'_{\min}$ including a single unit u from $t_{\min}$, $u \in t_{\min}$, $t'_{\min} \subseteq t_{\min}$.
4. For the unit u from $t'_{\min}$, find a unique candidate reservoir: $\{t'_1,...,t'_i,...,t'_{g-1}\}$ such that the distance between any unit from t'_i and any unit from t'_j satisfies (4.2-2). If no candidates meet (4.2-2), stop and return to step 3.
5. Perform NNWR: Match the unit in $t'_{\min}$ with up to m nearest units from each of other $g-1$ groups $t'_1,...,t'_i,...,t'_{g-1}$ by the overall distance D (see 4.2-3). Note here one unit from

the m nearest units may repeatedly be selected. Let $m_u^{'}$ be the actual number of units selected from $\mathrm{t}_i^{'}$ and matched to unit u, and $\lambda_{u,k}$ be the frequency that unit k is matched to unit u. The individual-level weight of unit k in response to unit u is then equal to $\frac{\lambda_{u,k}}{m_u^{'}}$.

6. Remove $\mathrm{t}_{\min}^{'}$ from $\mathrm{t}_{\min}$. Save the matched set: M_u and save $\frac{\lambda_{u,k}}{m_u^{'}}$ for the calculation of sample-level weight.
7. Repeat step 3 to 6 until all units from $\mathrm{t}_{\min}$ has been used.
8. Aggregate all matched sets into treatment groups. Calculate the sample-level weight of unit k: w_k by taking the sum of $\frac{\lambda_{u,k}}{m_u^{'}}$ over $\{\mathrm{M}_u, u \in \mathrm{t}_{\min}\}$ that include unit k. That is

$$w_k = \sum_{k \in \{\mathrm{M}_u, u \in \mathrm{t}_{\min}\}} \frac{\lambda_{u,k}}{m_u^{'}} \tag{4.2-4}$$

4.2.4 NNCV vs. NNCVWR

The first four steps are the same for NNCV and NNCVWR, which deal with filtering candidates using the new caliper width. Note both methods do not require the use of common support, because using caliper width already handles the overlap issue pretty well. Their difference starts from step 5 regarding the difference between NN and NNWR. Three factors determine the matching quality of NNCV and NNCVWR. The first factor is the caliper width defined in (4.2-1). Clearly, the greater the caliper width is the larger size of $\{\mathrm{t}_1^{'}, ..., \mathrm{t}_i^{'}, ..., \mathrm{t}_{g-1}^{'}\}$ will be. As the caliper increases to infinity, any d_{ij} satisfies (4.2-2), and caliper width matching is

equivalent to matching without caliper width. Therefore, when $m=1$, NNVC converges to NN. When $m>1$, NNVC converges to NN variable matching. The argument above proves the following proposition.

Proposition 4.1: The non-bipartite nearest neighbor matching is a particular case of non-bipartite nearest neighbor caliper width variable matching.

The second factor matching with or without replacement. In chapter 3 I have shown that NN tended to result in lower MSE of ACIE and ACME estimates than NNWR when using $\underbrace{1:1,...,:1:1}_{g}$ fixed number matching. However, NNWR had lower RB in certain cases. Whether NN is better than NNWR when using variable matching needs further exploration.

The third factor is the threshold m that controls the frequency of using NN or NNWR within $\{t_1^{'},...,t_i^{'},...,t_{g-1}^{'}\}$. In other words, this algorithm is a $\underbrace{1:m_u^{'},...,m_u^{'}:m_u^{'}}_{g}$ variable ratio matching, where $m^{'}$ is a constant and $0<=m_u^{'}<=m$, and $u\in t_{\min}$. The size of matched set is $(g-1)m^{'}$, and the matched sample size is $(g-1)\sum_{u\in t_{\min}} m_u^{'}$. Therefore, the matched sample size is a function of $m_u^{'}$. Since $m_u^{'}$ is controlled by m, w, and the choice between match with and without replacement, therefore these three factors affect both bias and variance of ACE. To determine an optimal value of w, m, and the choice between NN and NNWR, we conduct a Monte Carlo simulation study using $m\in\{1,2,3,4,5\}$ given two binary treatment variables. After identifying the optimal caliper width and m, we compare the optimal caliper width algorithm to the NN, NNWR, GIPTW and 3x3 SP under various treatment allocation settings.

4.3 Monte Carlo Simulations

4.3.1 Simulation Settings

The aim of the simulation study has two parts. In the first part, we examine the optimal caliper width and the optimal m of NNCV and NNCVWR given the w in formula (4.2-1) is between 0.2 to 3 by a grid of 0.05, and $m \in \{1,2,3,4,5\}$. We report RB and MSE of ACIE and ACME estimates. We choose the MSE of ACIE estimated from weighted ANOVA model as a performance indicator because of 3 reasons. See section 4.2.2 and 4.2.3 for the computation of weight for each subject. First, the choice of w, m, and the choice between NN and NNWR reveals an implicit trade-off between variance and bias. MSE is the sum of variance and squared bias. Therefore, concentrating on MSE allows researchers to identify the optimal solutions on the variance and bias trade-off. Second, we use the MSE of ACIE estimates because it reflects how well the distribution of baseline characteristics is balanced between treatment groups. Third, we select ANOVA rather linear regression to avoid the reduction in MSE due to covariate adjustment. In the second part, we compare the optimal NNCV and NNCVWR to NN, NNWR, GIPTW, and 3x3 SP as discussed in chapter 3 using MSE and RB of ACIE and ACME estimates. The data generation process and outcome stage analysis of NN, NNWR, GIPTW, and 3x3 SP is referred to section 3.4. When using NNCV and NNCVWR, we employ weighted linear regression controlling for all covariates related to outcome. The weight calculation on NNCV and NNCVWR are correspondingly referred to section 4.2.2 and 4.2.3.

4.3.2 Simulation Results Part I

Figure 4.1 to 4.8 present the relationship between w, m, matching with or without replacement, and MSE of ACIE in 31 prevalence of treatments. Within each figure, the panels on

the left side are corresponding to the NNCVWR, and the panels on the right side are regarding the NNCV. Each two panels in the same row are on the same prevalence case. Table 4.1 to 4.8 summarize the optimal w of a caliper width and the corresponding minimized MSE of ACIE estimates by methods, m and 31 prevalence of treatments.

Case: Prevalence of T_1 and T_2 are 0.2,0.2; 0.8,0.8 (Figure 4.1 and Table 4.1)

The patterns of case 0.2,0.2 and 0.8,0.8 are quite similar. When using NNCVWR, the MSE of ACIE estimates substantially decreased as m increased from 1 to 2, and slightly decreased as m rose from 2 to 5. The MSE decreased as w increased to 1, and increased as w increased to 3. Note the discrepancy between adjacent curves gradually reduced as m increased from 1 to 5. In the case of 0.2,0.2, the use of $w=1, m=5$ minimized MSE. In the case of 0.8,0.8, the use of $w=0.8, m=5$ minimized MSE.

When using NNCV, the use of m equal to 3 tended to result in lower MSE than the one with m equal to 1, and negligibly less MSE than those with other choices of m. In the case of 0.2,0.2, the use of $w=1.2, m=3$ minimized MSE. In the case of 0.8,0.8, the use of $w=1.3, m=3$ minimized MSE.

If we compare the NNCVWR and NNCV with the same m and optimal w. NNCV tended to result in lower MSE than NNCVWR when m equal to 1, and 2. However, NNCVWR is superior to NNCV when m is between to 3 to 5 and the prevalence of treatments is 0.2,0.2.

Case: Prevalence of T_1 and T_2 are 0.3,0.3; 0.7,0.7 (Figure 4.1 and Table 4.1)

The patterns of case 0.3,0.3 and 0.7,0.7 are pretty comparable.

When using NNCVWR, the MSE decreased as m increased from 1 to 5. Note the disparity between adjacent curves gradually reduced as m increased. In the case of 0.3,0.3, the use of

$w = 0.75, m = 5$ minimized MSE. In the case of 0.7,0.7, the use of $w = 0.9, m = 5$ minimized MSE.

When using NNCV, the use of m equal to 2 tended to result in negligible smaller MSE than those with other choices of m. In the case of 0.3,0.3, the use of $w = 1.15, m = 2$ minimized MSE. In the case of 0.7,0.7, the use of $w = 1, m = 2$ minimized MSE.

If we compare the NNCVWR and NNCV with optimal w the same m. NNCV tended to result in lower MSE than NNCVWR when m equal to 1. However, NNCVWR is superior to NNCV when m is between to 2 to 5.

Table 4.1 Summary of Optimal w, on Methods in the Cases of 0.2,0.2; 0.8,0.8; 0.3,0.3; and 0.7,0.7

Prevalence of T_1, T_2	Method	m	Optimal w	Lowest MSE of ACIE Estimates
0.2,0.2	NNCVWR/NNCV	1	1/1.25	1.61/1.24
0.2,0.2	NNCVWR/NNCV	2	1/1.25	1.04/0.99
0.2,0.2	NNCVWR/NNCV	3	1/1.2	0.86/0.99
0.2,0.2	NNCVWR/NNCV	4	1/1.2	0.81/1.06
0.2,0.2	NNCVWR/NNCV	5	1/1.25	0.76/1.10
0.8,0.8	NNCVWR/NNCV	1	0.85/1.4	1.58/1.25
0.8,0.8	NNCVWR/NNCV	2	0.7/1.1	0.98/1.02
0.8,0.8	NNCVWR/NNCV	3	0.8/1.3	0.86/1.02
0.8,0.8	NNCVWR/NNCV	4	0.8/1.15	0.79/1.05
0.8,0.8	NNCVWR/NNCV	5	0.8/1.15	0.75/1.07
0.3,0.3	NNCVWR/NNCV	1	1.05/1.15	0.95/0.63
0.3,0.3	NNCVWR/NNCV	2	0.85/1.15	0.51/0.57
0.3,0.3	NNCVWR/NNCV	3	0.75/1.25	0.45/0.59
0.3,0.3	NNCVWR/NNCV	4	0.75/1.25	0.41/0.66
0.3,0.3	NNCVWR/NNCV	5	0.75/1.25	0.39/0.70
0.7,0.7	NNCVWR/NNCV	1	0.6/1.1	0.90/0.65
0.7,0.7	NNCVWR/NNCV	2	0.75/1	0.52/0.60
0.7,0.7	NNCVWR/NNCV	3	0.8/1.2	0.45/0.66
0.7,0.7	NNCVWR/NNCV	4	0.9/1.35	0.42/0.70
0.7,0.7	NNCVWR/NNCV	5	0.9/1.35	0.40/0.75

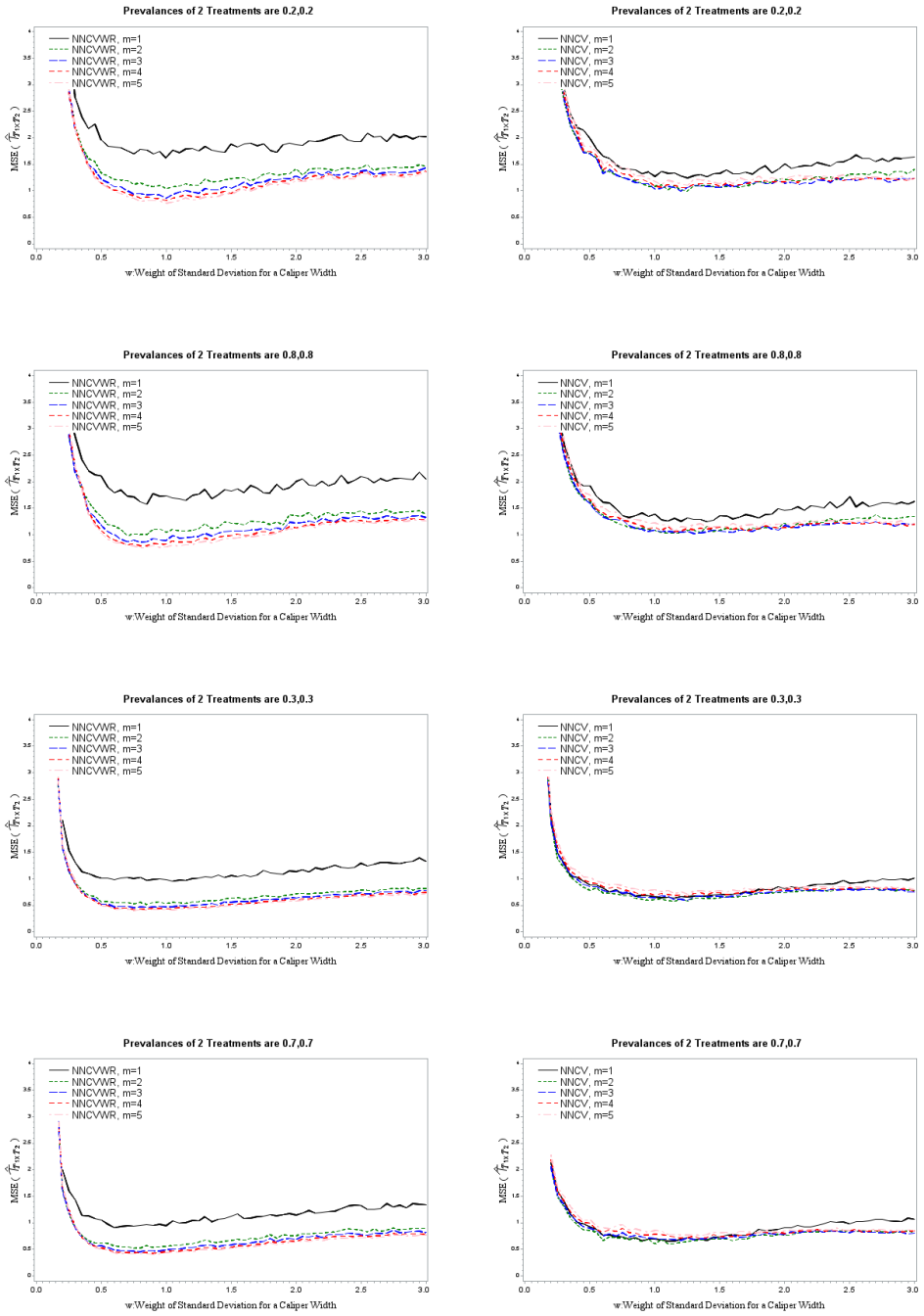

Figure 4.1 Caliper width, m vs. MSE of ACIE Part 1

Case: Prevalence of T_1 and T_2 are 0.4,0.4; 0.6,0.6 (Figure 4.2 and Table 4.2)

The patterns of case 0.4,0.4 and 0.6,0.6 are quite similar. When using NNCVWR, the MSE of ACIE estimates decreased as m increased from 1 to 5. The disparity between adjacent curves reduced as m rose from 1 to 5. In the case of 0.4,0.4, the use of $w = 0.65, m = 5$ minimized MSE. In the case of 0.6,0.6, the use of $w = 0.7, m = 5$ minimized MSE. When using NNCV, the use of m equal to 3 tended to result in lower MSE than the one with m equal to 1, and negligibly less MSE than those with other choices of m. In the case of 0.4,0.4, the use of $w = 1, m = 0.9$ and $w = 1, m = 2$ minimized MSE. In the case of 0.6,0.6, the use of $w = 1, m = 1$ and $w = 1.05, m = 2$ minimized MSE.

If we compare the NNCVWR and NNCV with the same m and optimal w. NNCV tended to produce lower MSE than NNCVWR when m equal to 1, and 2. However, NNCVWR outperformed NNCV when m is between to 3 to 5.

Case: Prevalence of T_1 and T_2 are 0.5,0.5 (Figure 4.2 and Table 4.2)

When using NNCVWR, the use of m equal to 1 is inferior to other choices of m. The MSE are pretty closed from one to another when m is between 2 to 5. The choice of w had a pretty small impact on MSE as m increased from around 0.3. The use of $w = 2.95, m = 1$ and $w = 3, m = 2$ minimized MSE.

When using NNCV, the choice of m had small effect on MSE. The use of $w = 2.95, m = 1$ and $w = 3, m = 2$ minimized MSE.

Table 4.2 Summary of Optimal w, on Methods in the Cases of 0.4,0.4; 0.6,0.6; and 0.5,0.5.

Prevalence of T_1, T_2	Method	m	Optimal w	Lowest MSE of ACIE Estimates
0.4,0.4	NNCVWR/NNCV	1	0.55/0.9	0.75/0.50
0.4,0.4	NNCVWR/NNCV	2	0.75/1	0.40/0.50
0.4,0.4	NNCVWR/NNCV	3	0.75/1.35	0.34/0.54
0.4,0.4	NNCVWR/NNCV	4	0.75/1.4	0.32/0.60
0.4,0.4	NNCVWR/NNCV	5	0.65/1.05	0.30/0.65
0.6,0.6	NNCVWR/NNCV	1	0.5/1	0.77/0.48
0.6,0.6	NNCVWR/NNCV	2	0.65/1.05	0.38/0.48
0.6,0.6	NNCVWR/NNCV	3	0.65/1.15	0.34/0.53
0.6,0.6	NNCVWR/NNCV	4	0.7/1.35	0.32/0.56
0.6,0.6	NNCVWR/NNCV	5	0.65/1.25	0.30/0.63
0.5,0.5	NNCVWR/NNCV	1	1.2/2.95	0.67/0.31
0.5,0.5	NNCVWR/NNCV	2	0.75/3	0.32/0.31
0.5,0.5	NNCVWR/NNCV	3	0.65/3	0.29/0.35
0.5,0.5	NNCVWR/NNCV	4	0.65/3	0.27/0.37
0.5,0.5	NNCVWR/NNCV	5	0.65/2.95	0.26/0.41

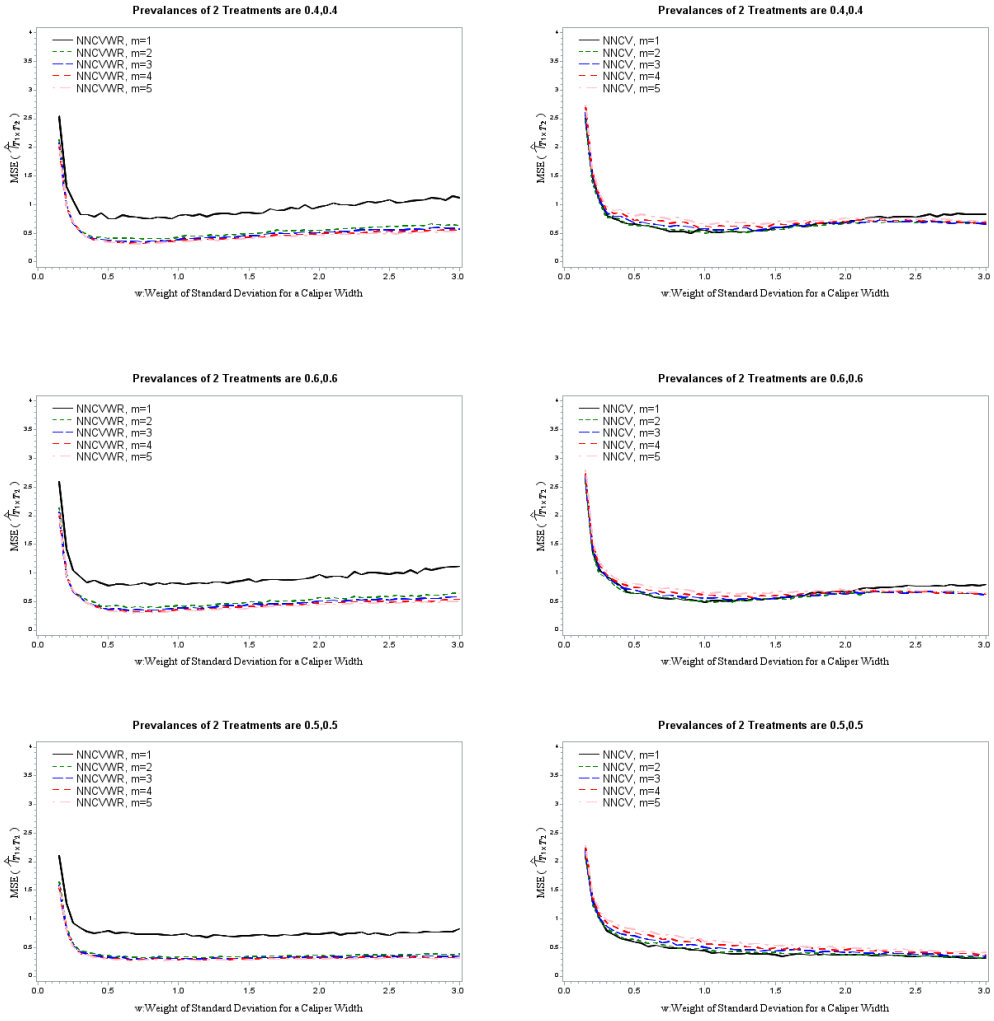

Figure 4.2 Caliper width, m vs. MSE of ACIE Part 2

Case: Prevalence of T_1 and T_2 are 0.2,0.3; 0.3,0.2; 0.2,0.4; 0.4,0.2 (Figure 4.3 and Table 4.3)

The patterns of case 0.2,0.3; 0.3,0.2; 0.2,0.4; and 0.4,0.2 are fairly comparable. When using NNCVWR, the MSE of ACIE estimates decreased as m increased from 1 to 5. The dissimilarity between adjacent curves diminished as m increased. In the case of 0.2,0.3, the use of $w = 0.8, m = 5$ minimized MSE. In the case of 0.3,0.2, the use of $w = 1.1, m = 5$ minimized MSE. In the case of 0.2,0.4, the use of $w = 0.9, m = 5$ minimized MSE. In the case of 0.4,0.2, the use of $w = 0.75, m = 5$ minimized MSE.

When using NNCV, the change of m had minor impact on MSE. In the case of 0.2,0.3, the use of $w=1.25, m=2$ minimized MSE. In the case of 0.3,0.2, the use of $w=1.4, m=2$ and minimized MSE. In the case of 0.2,0.4, the use of $w=1.35, m=2$ minimized MSE. In the case of 0.4,0.2, the use of $w=1.2, m=2$ minimized MSE.

If we compare the NNCVWR and NNCV with the same m and optimal w. NNCV tended to produce lower MSE than NNCVWR when m equal to 1. However, NNCVWR is better than NNCV when m is between to 2 to 5.

Table 4.3 Summary of Optimal w, on Methods in the Cases of 0.2,0.3; 0.3,0.2; 0.2,0.4; and 0.4,0.2

Prevalence of T_1, T_2	Method	m	Optimal w	Lowest MSE of ACIE Estimates
0.2, 0.3	NNCVWR/NNCV	1	1/1.4	1.26/0.89
0.2, 0.3	NNCVWR/NNCV	2	0.85/1.25	0.73/0.79
0.2, 0.3	NNCVWR/NNCV	3	0.85/1.3	0.64/0.83
0.2, 0.3	NNCVWR/NNCV	4	0.95/1.25	0.59/0.86
0.2, 0.3	NNCVWR/NNCV	5	0.8/1.25	0.56/0.92
0.3, 0.2	NNCVWR/NNCV	1	0.9/1.1	1.25/0.91
0.3, 0.2	NNCVWR/NNCV	2	1.1/1.4	0.74/0.83
0.3, 0.2	NNCVWR/NNCV	3	1.1/1.4	0.66/0.86
0.3, 0.2	NNCVWR/NNCV	4	0.95/1.4	0.62/0.90
0.3, 0.2	NNCVWR/NNCV	5	1.1/1.35	0.58/0.90
0.2, 0.4	NNCVWR/NNCV	1	0.6/1.1	1.15/0.78
0.2, 0.4	NNCVWR/NNCV	2	0.9/1.35	0.63/0.72
0.2, 0.4	NNCVWR/NNCV	3	0.9/1.35	0.54/0.76
0.2, 0.4	NNCVWR/NNCV	4	0.95/0.9	0.50/0.80
0.2, 0.4	NNCVWR/NNCV	5	0.9/0.9	0.47/0.87
0.4, 0.2	NNCVWR/NNCV	1	0.6/1.2	1.15/0.76
0.4, 0.2	NNCVWR/NNCV	2	0.6/1.2	0.62/0.75
0.4, 0.2	NNCVWR/NNCV	3	1/1.15	0.57/0.77
0.4, 0.2	NNCVWR/NNCV	4	0.95/1.15	0.52/0.84
0.4, 0.2	NNCVWR/NNCV	5	0.75/1.25	0.49/0.90

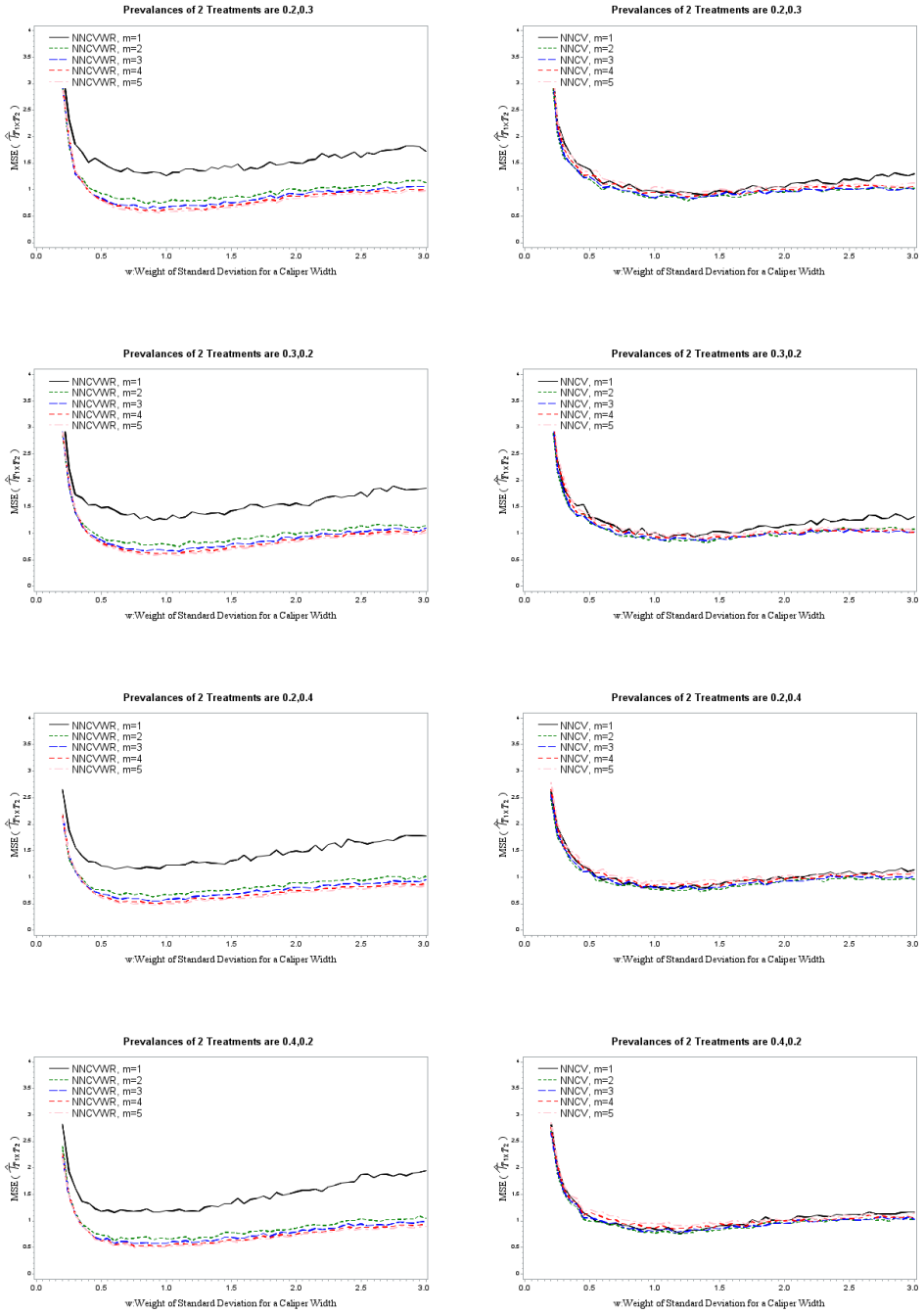

Figure 4.3 Caliper width, m vs. MSE of ACIE Part 3

Case: Prevalence of T_1 and T_2 are 0.2,0.5; 0.5,0.2; 0.3,0.4; 0.4,0.3 (Figure 4.4 and Table 4.4)

The patterns on Figure 4.4 is pretty similar to Figure 4.3

Using NNCVWR, the choices of $w=0.95, m=5$, $w=1.05, m=5$, $w=0.6, m=5$, and $w=0.8, m=5$ respectively minimized MSE of ACIE given the cases of 0.2,0.5; 0.5,0.2; 0.3,0.4; and 0.4,0.3.

When using NNCV, the choice of $w=1.4, m=2$, $w=1.1, m=2$, $w=1.15, m=2$, and $w=1.15, m=2$ respectively minimized MSE of ACIE given the cases of 0.2,0.5; 0.5,0.2; 0.3,0.4; and 0.4,0.3.

Table 4.4 Summary of Optimal w, on Methods in the Cases of 0.2,0.5; 0.5,0.2; 0.3,0.4; and 0.4,0.3

Prevalence of T_1, T_2	Method	m	Optimal w	Lowest MSE of ACIE Estimates
0.2, 0.5	NNCVWR/NNCV	1	0.85/1.05	1.05/0.72
0.2, 0.5	NNCVWR/NNCV	2	0.8/1.4	0.61/0.74
0.2, 0.5	NNCVWR/NNCV	3	0.8/1.1	0.54/0.81
0.2, 0.5	NNCVWR/NNCV	4	0.95/1	0.52/0.89
0.2, 0.5	NNCVWR/NNCV	5	0.95/1.25	0.49/0.96
0.5, 0.2	NNCVWR/NNCV	1	0.75/1.25	0.99/0.67
0.5, 0.2	NNCVWR/NNCV	2	1.05/1.1	0.56/0.65
0.5, 0.2	NNCVWR/NNCV	3	1.05/1	0.51/0.75
0.5, 0.2	NNCVWR/NNCV	4	0.95/1.2	0.48/0.79
0.5, 0.2	NNCVWR/NNCV	5	1.05/1	0.46/0.87
0.3, 0.4	NNCVWR/NNCV	1	1.4/1.15	0.84/0.56
0.3, 0.4	NNCVWR/NNCV	2	0.7/1.15	0.45/0.57
0.3, 0.4	NNCVWR/NNCV	3	0.9/0.85	0.40/0.60
0.3, 0.4	NNCVWR/NNCV	4	0.85/1.1	0.38/0.65
0.3, 0.4	NNCVWR/NNCV	5	0.6/1.1	0.36/0.70
0.4, 0.3	NNCVWR/NNCV	1	1.1/1.05	0.83/0.58
0.4, 0.3	NNCVWR/NNCV	2	0.8/1.15	0.46/0.55
0.4, 0.3	NNCVWR/NNCV	3	0.65/0.95	0.41/0.60
0.4, 0.3	NNCVWR/NNCV	4	0.75/1.05	0.38/0.62
0.4, 0.3	NNCVWR/NNCV	5	0.8/1.25	0.36/0.66

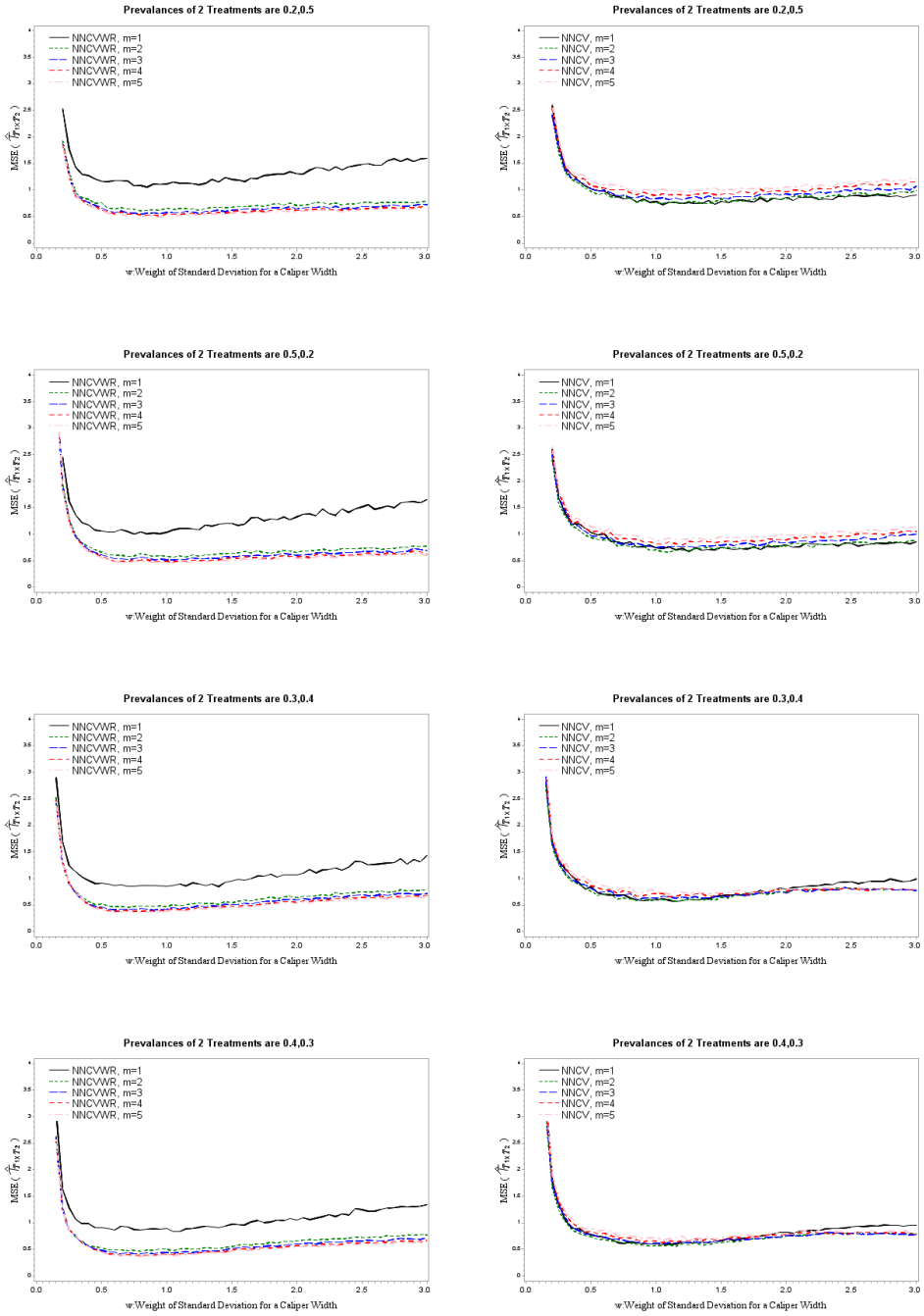

Figure 4.4 Caliper width, m vs. MSE of ACIE Part 4

Case: Prevalence of T_1 and T_2 are 0.2,0.6; 0.2,0.7; 0.2,0.8; 0.3,0.6 (Figure 4.5 and Table 4.5)

The patterns of case 0.2,0.6; 0.2,0.7; 0.2,0.8 and 0.3,0.6 are very close. When using NNCVWR, the MSE of ACIE estimates decreased as m increased from 1 to 5. The adjacent curves gradually coincided as m rose from 2 to 5. Each curve stabilized as w increased from 0.5 to 3. The choices of $w = 2.6, m = 5$, $w = 2.25, m = 5$, $w = 2.85, m = 5$, and $w = 3, m = 5$ respectively minimized MSE of ACIE given the cases of 0.2,0.6; 0.2,0.7; 0.2,0.8 and 0.3,0.6. When using NNCV, m had a negligible impact on MSE. The use of $w = 2.6, m = 5$, $w = 2.25, m = 5$, $w = 2.85, m = 5$, and $w = 3, m = 5$ respectively minimized MSE of ACIE given the cases of 0.2,0.6; 0.2,0.7; 0.2,0.8 and 0.3,0.6.

In the cases of 0.2,0.6, and 0.3,0.6 NNCV tended to produce lower MSE than NNCVWR when m equal to 1 and 2. Conversely, NNCVWR is better than NNCV when m is between to 2 to 5. In the case of 0.2,0.7, NNCV tended to result in lower MSE than NNCVWR when m equal to 1 and 2. Two methods had the same performances when m equal to 3. Otherwise, NNCVWR is better than NNCV when m is between to 3 to 5. In the case of 0.2,0.8, NNCV was superior to NNCVWR when m is less than 4, but worse than NNCVWR otherwise.

Table 4.5 Summary of Optimal w, on Methods in the Cases of 0.2,0.6; 0.2,0.7; 0.2,0.8; 0.3,0.6

Prevalence of T_1, T_2	Method	m	Optimal w	Lowest MSE of ACIE Estimates
0.2,0.6	NNCVWR/NNCV	1	1.05/1.65	1.12/0.58
0.2,0.6	NNCVWR/NNCV	2	2.6/1.85	0.58/0.57
0.2,0.6	NNCVWR/NNCV	3	2.7/1.85	0.51/0.62
0.2,0.6	NNCVWR/NNCV	4	2.6/1.85	0.48/0.66
0.2,0.6	NNCVWR/NNCV	5	2.6/2.25	0.46/0.75
0.2,0.7	NNCVWR/NNCV	1	1/2.9	1.13/0.62
0.2,0.7	NNCVWR/NNCV	2	2.65/2.95	0.58/0.48
0.2,0.7	NNCVWR/NNCV	3	2.8/2.5	0.52/0.52
0.2,0.7	NNCVWR/NNCV	4	2.25/2.5	0.48/0.60
0.2,0.7	NNCVWR/NNCV	5	2.25/2.7	0.45/0.65
0.2,0.8	NNCVWR/NNCV	1	2.6/2.65	1.37/0.94
0.2,0.8	NNCVWR/NNCV	2	2.55/2.65	0.83/0.65
0.2,0.8	NNCVWR/NNCV	3	2.2/3	0.71/0.58
0.2,0.8	NNCVWR/NNCV	4	2.2/3	0.64/0.60
0.2,0.8	NNCVWR/NNCV	5	2.85/2.9	0.61/0.66
0.3,0.6	NNCVWR/NNCV	1	1.65/2	0.74/0.37
0.3,0.6	NNCVWR/NNCV	2	2.4/2.45	0.38/0.36
0.3,0.6	NNCVWR/NNCV	3	1.55/2.25	0.34/0.42
0.3,0.6	NNCVWR/NNCV	4	2.4/2.2	0.32/0.46
0.3,0.6	NNCVWR/NNCV	5	3/2.85	0.31/0.53

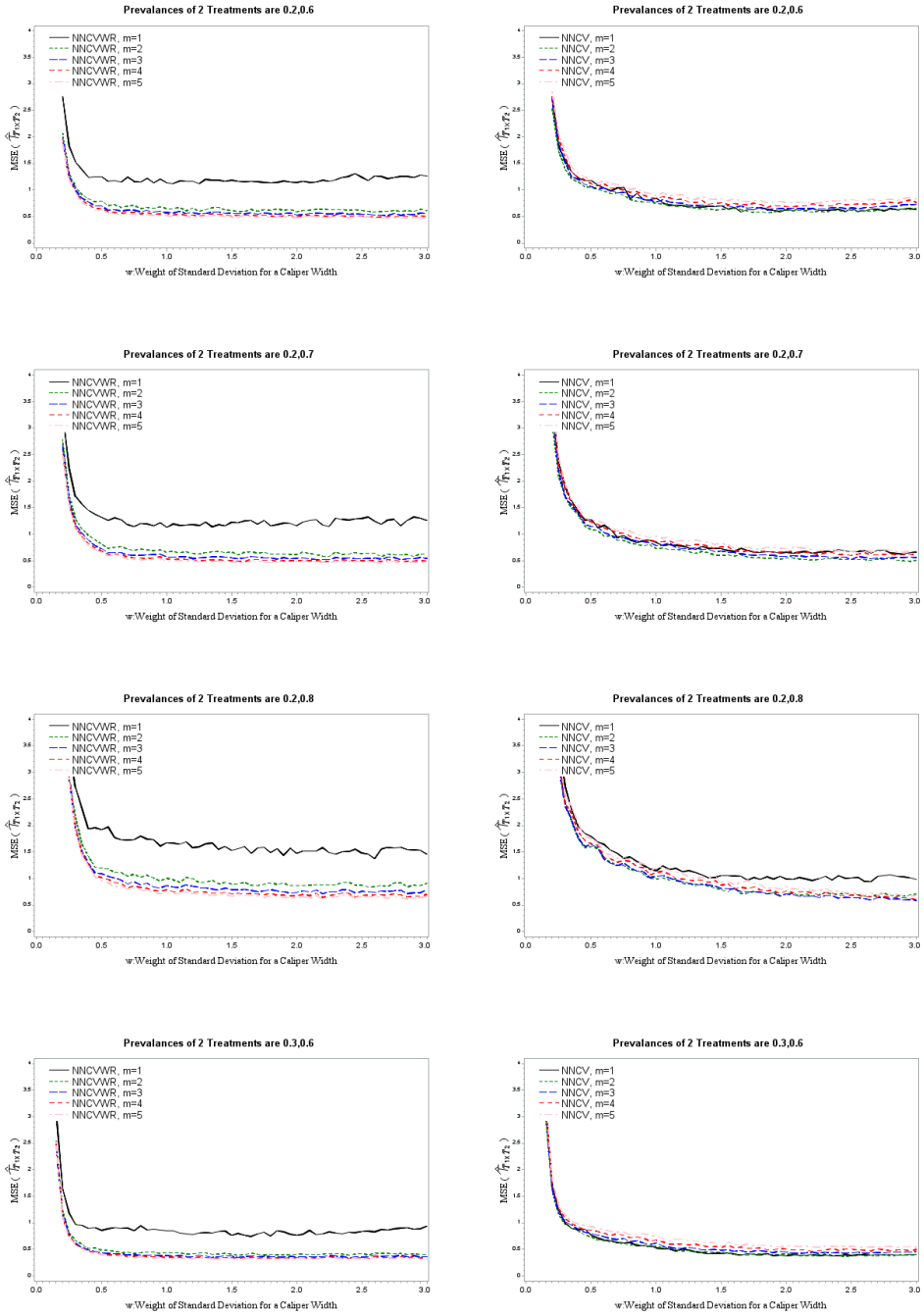

Figure 4.5 Caliper width, m vs. MSE of ACIE Part 5

Case: Prevalence of T_1 and T_2 are 0.3,0.5; 0.5,0.3; 0.4,0.5; 0.5,0.4 (Figure 4.6 and Table 4.6)

The patterns of case 0.3,0.5; 0.5,0.3; 0.4,0.5 and 0.5,0.4 are similar.

When using NNCVWR, the MSE of ACIE estimates decreased as m increased from 1 to 5. The adjacent curves tended to overlap as m rose from 2 to 5. Each curve stabilized as w increased from 0.45 to 3. The choices of $w=0.8, m=5$, $w=0.85, m=5$, $w=0.75, m=5$, and $w=0.9, m=5$ respectively minimized MSE of ACIE given the cases of 0.3,0.5; 0.5,0.3; 0.4,0.5 and 0.5,0.4. When using NNCV, m had a pretty low effect on MSE. The use of $w=1.3, m=1$, $w=1, m=0.6$, $w=1.5, m=1$, and $w=1.3, m=1$ respectively minimized MSE of ACIE given the cases of 0.3,0.5; 0.5,0.3; 0.4,0.5 and 0.5,0.4. NNCVWR is inferior to NNCV when m equal to 1, but better than NNCV otherwise.

Table 4.6 Summary of Optimal w, on Methods in the Cases of 0.3,0.5; 0.5,0.3; 0.4,0.5; 0.5,0.4

Prevalence of T_1, T_2	Method	m	Optimal w	Lowest MSE of ACIE Estimates
0.3,0.5	NNCVWR/NNCV	1	0.75/1.3	0.81/0.49
0.3,0.5	NNCVWR/NNCV	2	0.8/1.35	0.43/0.49
0.3,0.5	NNCVWR/NNCV	3	0.85/1.7	0.39/0.54
0.3,0.5	NNCVWR/NNCV	4	1/1.45	0.37/0.59
0.3,0.5	NNCVWR/NNCV	5	0.8/1.65	0.35/0.65
0.5,0.3	NNCVWR/NNCV	1	0.6/1.2	0.81/0.50
0.5,0.3	NNCVWR/NNCV	2	0.65/1.45	0.42/0.50
0.5,0.3	NNCVWR/NNCV	3	0.85/1.6	0.36/0.53
0.5,0.3	NNCVWR/NNCV	4	0.85/1.6	0.35/0.57
0.5,0.3	NNCVWR/NNCV	5	0.85/1.7	0.33/0.64
0.4,0.5	NNCVWR/NNCV	1	1.45/1.5	0.70/0.43
0.4,0.5	NNCVWR/NNCV	2	1.25/2.9	0.36/0.44
0.4,0.5	NNCVWR/NNCV	3	0.85/2.9	0.33/0.44
0.4,0.5	NNCVWR/NNCV	4	0.85/2.95	0.30/0.47
0.4,0.5	NNCVWR/NNCV	5	0.75/2.9	0.29/0.48
0.5,0.4	NNCVWR/NNCV	1	1.45/1.3	0.67/0.42
0.5,0.4	NNCVWR/NNCV	2	0.8/2.3	0.32/0.44
0.5,0.4	NNCVWR/NNCV	3	0.85/2.55	0.30/0.46
0.5,0.4	NNCVWR/NNCV	4	0.8/3	0.28/0.49
0.5,0.4	NNCVWR/NNCV	5	0.9/3	0.27/0.52

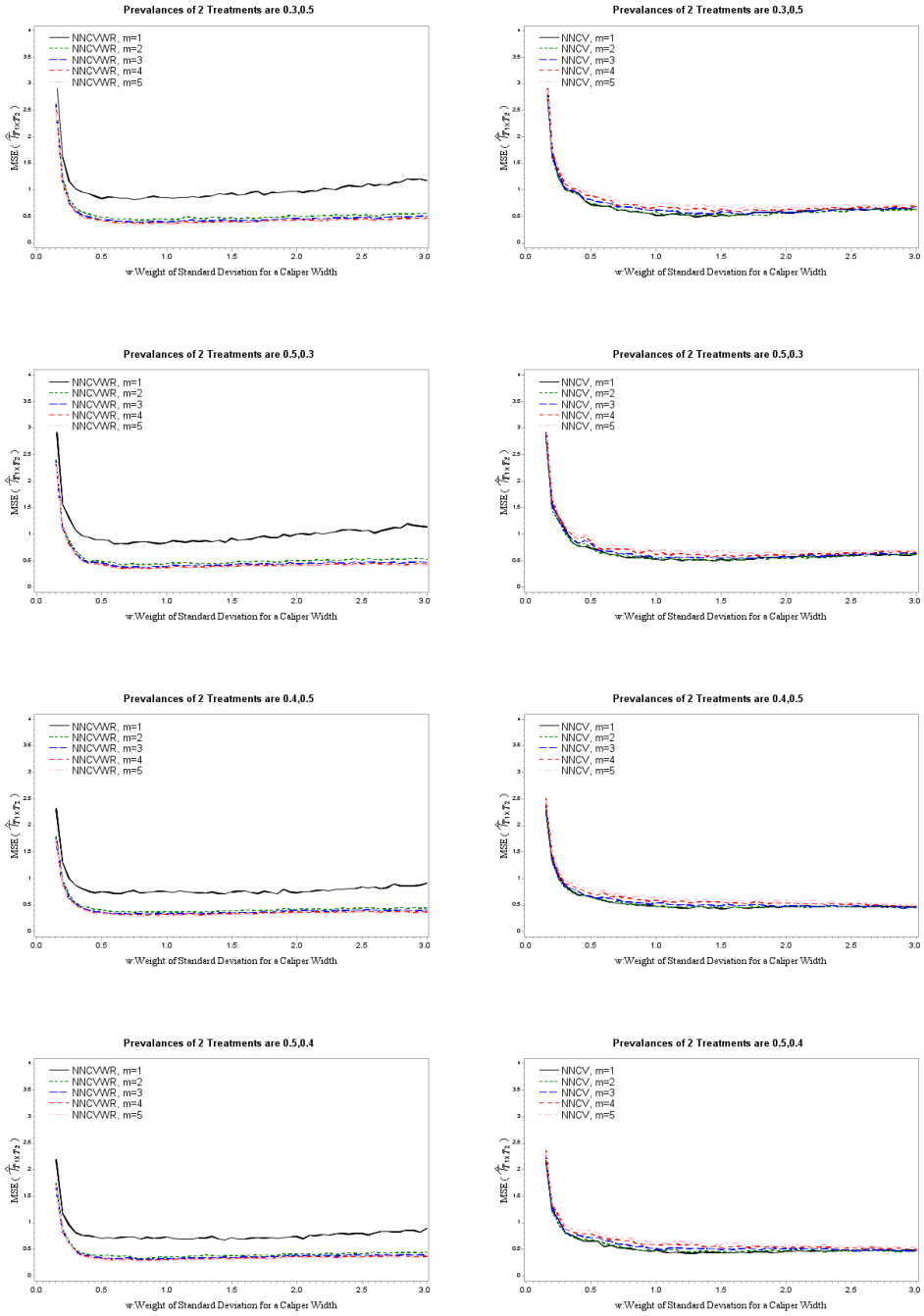

Figure 4.6 Caliper width, m vs. MSE of ACIE Part 6

Case: Prevalence of T_1 and T_2 are 0.3,0.7; 0.3,0.8; 0.4,0.6; 0.4,0.7 (Figure 4.7 and Table 4.7)

The patterns of case 0.3,0.7; 0.3,0.8; 0.4,0.6 and 0.4,0.7 are comparable.

When using NNCVWR, the MSE decreased as m increased from 1 to 5. The difference between adjacent curves diminished as m rose from 2 to 5. Each curve stabilized as w increased from 0.5 to 3. The choices of $w = 2.15, m = 5$, $w = 1.9, m = 5$, $w = 2.85, m = 5$, and $w = 2.75, m = 5$ respectively minimized MSE of ACIE given the cases of 0.3,0.7; 0.3,0.8; 0.4,0.6 and 0.4,0.7.

When using NNCV, m had a small influence on MSE. The use of $w = 2.95, m = 2$, $w = 2.5, m = 2$, $w = 2.85, m = 1$, and $w = 2.45, m = 2$ respectively minimized MSE of ACIE given the cases of 0.3,0.7; 0.3,0.8; 0.4,0.6 and 0.4,0.7. NNCVWR is inferior to NNCV when m equal to 1 and 2, but better than NNCV when m is between 3 and 5.

Table 4.7 Summary of Optimal w, on Methods in the Cases of 0.3,0.7; 0.3,0.8; 0.4,0.6; 0.4,0.7

Prevalence of T_1, T_2	Method	m	Optimal w	Lowest MSE of ACIE Estimates
0.3,0.7	NNCVWR/NNCV	1	2.9/2.3	0.82/0.45
0.3,0.7	NNCVWR/NNCV	2	2.05/2.95	0.44/0.35
0.3,0.7	NNCVWR/NNCV	3	1.65/2.85	0.39/0.41
0.3,0.7	NNCVWR/NNCV	4	2.15/2.75	0.37/0.46
0.3,0.7	NNCVWR/NNCV	5	2.15/3	0.35/0.52
0.3,0.8	NNCVWR/NNCV	1	1.25/2.95	1.21/0.68
0.3,0.8	NNCVWR/NNCV	2	2.5/2.5	0.66/0.54
0.3,0.8	NNCVWR/NNCV	3	1.9/2.5	0.58/0.57
0.3,0.8	NNCVWR/NNCV	4	1.9/2	0.53/0.64
0.3,0.8	NNCVWR/NNCV	5	1.9/2.45	0.50/0.68
0.4,0.6	NNCVWR/NNCV	1	2.7/2.85	0.60/0.26
0.4,0.6	NNCVWR/NNCV	2	2.7/2.85	0.31/0.27
0.4,0.6	NNCVWR/NNCV	3	2.85/2.85	0.28/0.31
0.4,0.6	NNCVWR/NNCV	4	2.85/3	0.26/0.34
0.4,0.6	NNCVWR/NNCV	5	2.85/2.9	0.25/0.38
0.4,0.7	NNCVWR/NNCV	1	1.95/2.15	0.73/0.37
0.4,0.7	NNCVWR/NNCV	2	2/2.45	0.37/0.35
0.4,0.7	NNCVWR/NNCV	3	2.35/2.9	0.34/0.40
0.4,0.7	NNCVWR/NNCV	4	2.45/2.8	0.31/0.43
0.4,0.7	NNCVWR/NNCV	5	2.75/2.65	0.30/0.48

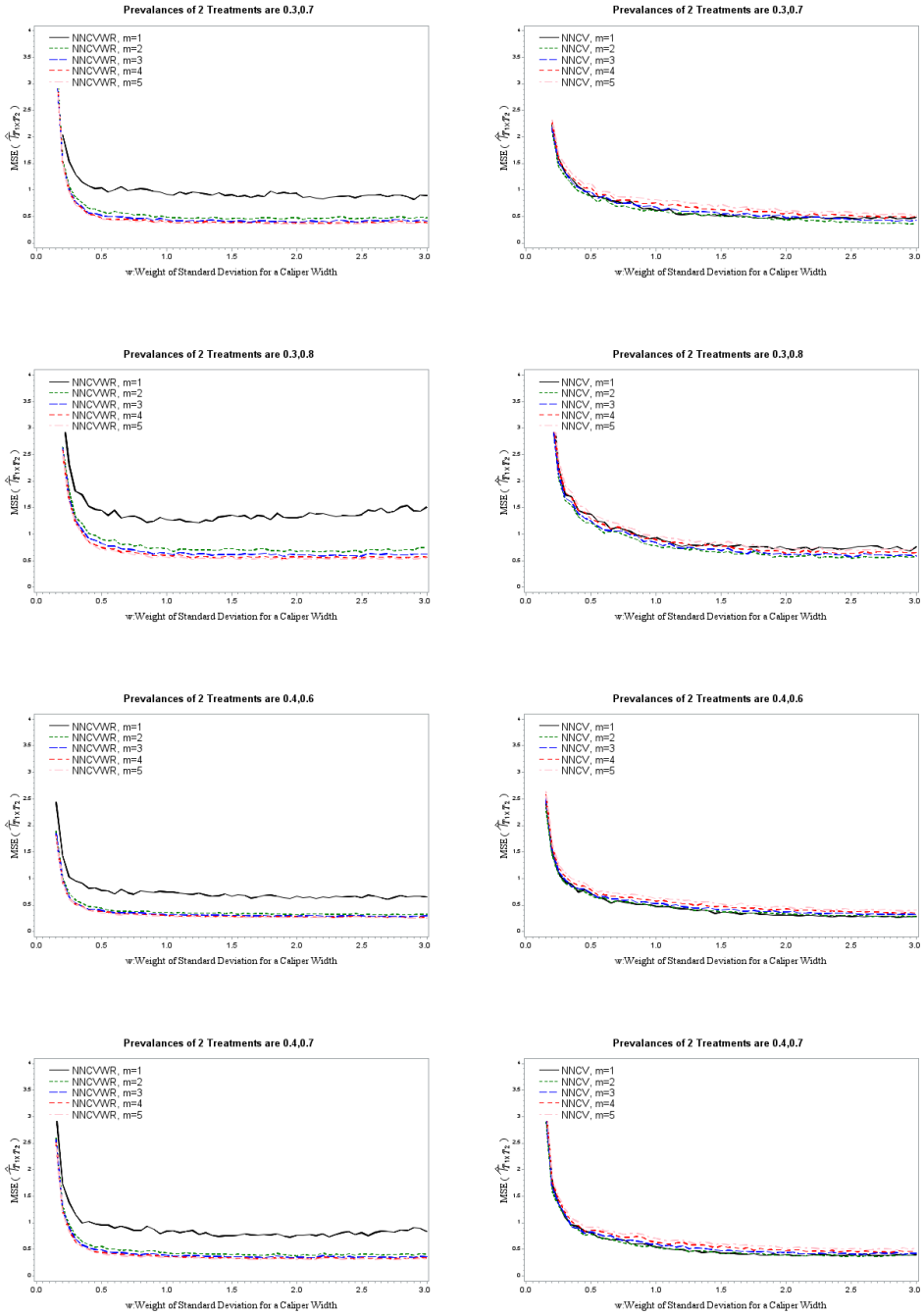

Figure 4.7 Caliper width, m vs. MSE of ACIE Part 7

Case: Prevalence of T_1 and T_2 are 0.4,0.8; 0.5,0.6; 0.5,0.7; 0.5,0.8 (Figure 4.8 and Table 4.8)

The patterns of case 0.4,0.8; 0.5,0.6; 0.5,0.7 and 0.5,0.8 are similar.

When using NNCVWR, the MSE decreased as m increased from 1 to 5. The difference between adjacent curves reduced as m increased. Each curve stabilized as w rose from 0.5 to 3. The choices of $w=2.35, m=5$, $w=0.9, m=5$, $w=1, m=5$, and $w=0.8, m=5$ respectively minimized MSE of ACIE given the cases of 0.4,0.8; 0.5,0.6; 0.5,0.7 and 0.5,0.8. When using NNCV, m had little impact on MSE. The use of $w=2, m=2$, $w=1.6, m=2$, $w=1.35, m=1$, and $w=1.4, m=2$ respectively minimized MSE of ACIE given the cases of 0.4,0.8; 0.5,0.6; 0.5,0.7 and 0.5,0.8. NNCVWR outperformed NNCV when m is between 2 and 5 in the cases of 0.5,0.6; 0.5,0.7 and 0.5,0.8, as well as when m is between 3 and 5 in the cases of 0.4,0.8.

Table 4.8 Summary of Optimal w, on Methods in the Cases of 0.4,0.8; 0.5,0.6; 0.5,0.7; 0.5,0.8

Prevalence of T_1, T_2	Method	m	Optimal w	Lowest MSE of ACIE Estimates
0.4,0.8	NNCVWR/NNCV	1	1.25/1.85	1.14/0.58
0.4,0.8	NNCVWR/NNCV	2	2.35/2	0.57/0.56
0.4,0.8	NNCVWR/NNCV	3	2.5/1.75	0.52/0.63
0.4,0.8	NNCVWR/NNCV	4	2.35/2.2	0.48/0.68
0.4,0.8	NNCVWR/NNCV	5	2.35/1.75	0.47/0.75
0.5,0.6	NNCVWR/NNCV	1	1.45/1.15	0.70/0.40
0.5,0.6	NNCVWR/NNCV	2	0.8/1.6	0.34/0.40
0.5,0.6	NNCVWR/NNCV	3	0.8/1.45	0.31/0.44
0.5,0.6	NNCVWR/NNCV	4	0.8/3	0.29/0.48
0.5,0.6	NNCVWR/NNCV	5	0.9/2.55	0.28/0.51
0.5,0.7	NNCVWR/NNCV	1	0.85/1.35	0.78/0.50
0.5,0.7	NNCVWR/NNCV	2	1/1.55	0.42/0.52
0.5,0.7	NNCVWR/NNCV	3	0.75/1.2	0.37/0.56
0.5,0.7	NNCVWR/NNCV	4	1/1.7	0.35/0.61
0.5,0.7	NNCVWR/NNCV	5	1/1.6	0.33/0.66
0.5,0.8	NNCVWR/NNCV	1	0.7/1.35	1.02/0.67
0.5,0.8	NNCVWR/NNCV	2	0.95/1.4	0.59/0.64
0.5,0.8	NNCVWR/NNCV	3	0.95/1.35	0.52/0.70
0.5,0.8	NNCVWR/NNCV	4	0.95/1.6	0.49/0.76
0.5,0.8	NNCVWR/NNCV	5	0.8/1.35	0.46/0.83

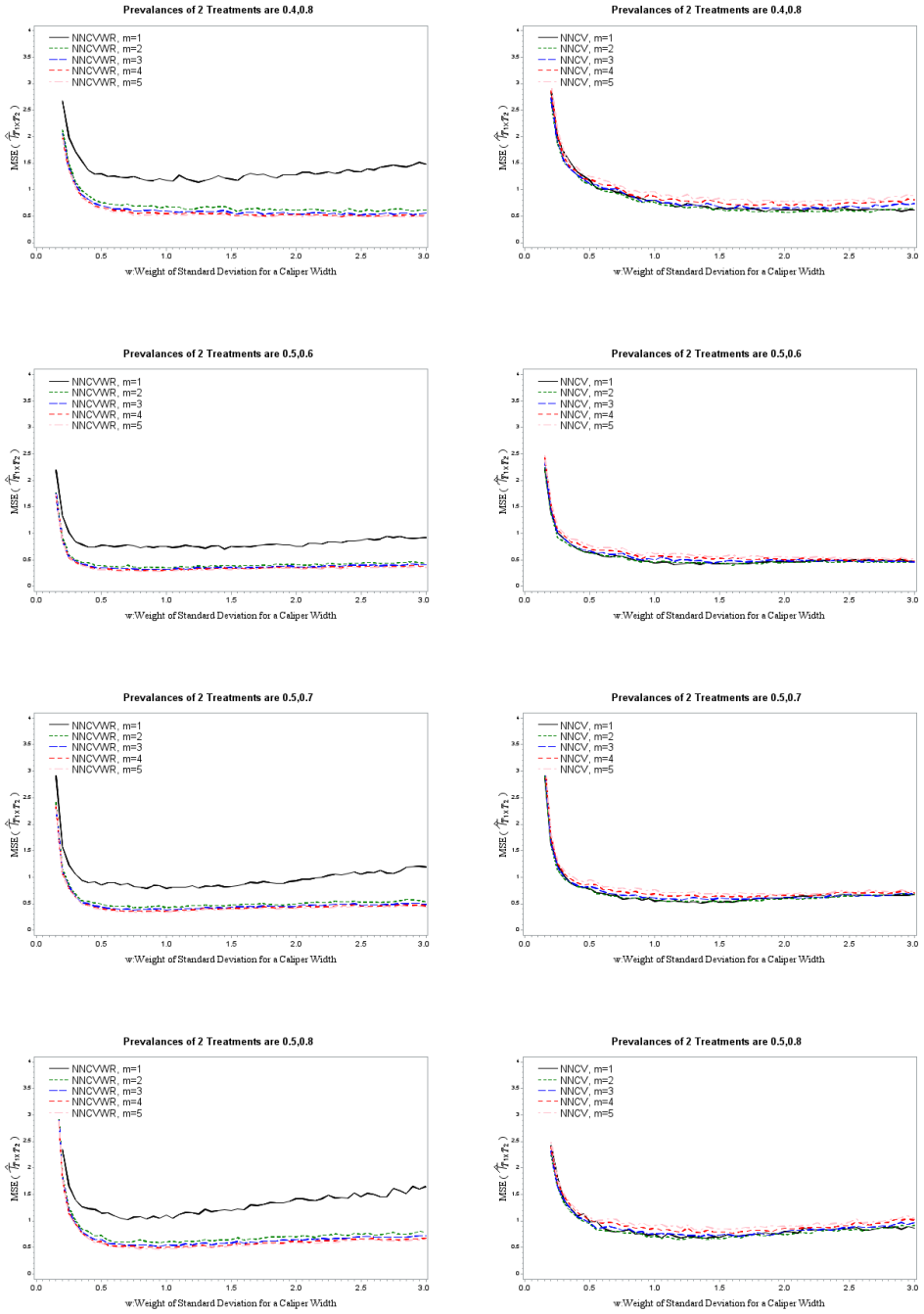

Figure 4.8 Caliper width, m vs. MSE of ACIE Part 8

Table 4.9 compares the NNCVWR and NNCV using their optimal m and w corresponding to minimum MSE of ACIE Estimates. In general, when using NNCVWR, the optimal m is always 5, and the optimal w is between 0.6 and 1 in 71% prevalence of treatments scenarios and is

between 1.9 to 3 in 29% prevalence of treatments scenarios. When using NNCV, the optimal m is 3 in 58% cases, 2 in 32% cases, and 3 in 10% cases. The optimal w ranged from 0.9 to 3. The optimal NNCVWR is superior to the optimal NNCV in 30 out of 32 cases (93.75%), the same as optimal NNCV in 1 case, and inferior to NNCV in 1 case.

Table 4.9 Summary of Optimal m and w of NNCVWR and NNCV Based on the Minimum MSE of ACIE Estimates by Prevalence of T_1, T_2

Prevalence of T_1, T_2	Method	Optimal m	Optimal w	Lowest MSE of ACIE Estimates
0.2,0.2	NNCVWR/NNCV	5/3	1/1.2	0.76/0.99
0.2,0.3	NNCVWR/NNCV	5/2	0.8/1.25	0.56/0.79
0.2,0.4	NNCVWR/NNCV	5/2	0.9/1.35	0.47/0.72
0.2,0.5	NNCVWR/NNCV	5/1	0.95/1.05	0.49/0.72
0.2,0.6	NNCVWR/NNCV	5/2	2.6/1.85	0.46/0.57
0.2,0.7	NNCVWR/NNCV	5/2	2.25/2.95	0.45/0.48
0.2,0.8	NNCVWR/NNCV	5/3	2.85/3	0.61/0.58
0.3,0.2	NNCVWR/NNCV	5/2	1.1/1.4	0.58/0.83
0.3,0.3	NNCVWR/NNCV	5/2	0.75/1.15	0.39/0.57
0.3,0.4	NNCVWR/NNCV	5/1	0.6/1.15	0.36/0.56
0.3,0.5	NNCVWR/NNCV	5/1	0.8/1.3	0.35/0.49
0.3,0.6	NNCVWR/NNCV	5/2	3/2.45	0.31/0.36
0.3,0.7	NNCVWR/NNCV	5/2	2.15/2.95	0.35/0.35
0.3,0.8	NNCVWR/NNCV	5/2	1.9/2.5	0.50/0.54
0.4,0.2	NNCVWR/NNCV	5/2	0.75/1.2	0.49/0.75
0.4,0.3	NNCVWR/NNCV	5/2	0.8/1.15	0.36/0.55
0.4,0.4	NNCVWR/NNCV	5/1	0.65/0.9	0.30/0.50
0.4,0.5	NNCVWR/NNCV	5/1	0.75/1.5	0.29/0.43
0.4,0.6	NNCVWR/NNCV	5/1	2.85/2.85	0.25/0.26
0.4,0.7	NNCVWR/NNCV	5/2	2.75/2.45	0.30/0.35
0.4,0.8	NNCVWR/NNCV	5/2	2.35/2	0.47/0.56
0.5,0.2	NNCVWR/NNCV	5/2	1.05/1.1	0.46/0.65
0.5,0.3	NNCVWR/NNCV	5/1	0.85/1.2	0.33/0.50
0.5,0.4	NNCVWR/NNCV	5/1	0.9/1.3	0.27/0.42
0.5,0.5	NNCVWR/NNCV	5/1	0.65/2.95	0.26/0.31
0.5,0.6	NNCVWR/NNCV	5/2	0.9/1.6	0.28/0.40
0.5,0.7	NNCVWR/NNCV	5/1	1/1.35	0.33/0.50
0.5,0.8	NNCVWR/NNCV	5/2	0.8/1.4	0.46/0.64
0.6,0.6	NNCVWR/NNCV	5/2	0.65/1.05	0.30/0.48
0.7,0.7	NNCVWR/NNCV	5/2	0.9/1	0.40/0.60
0.8,0.8	NNCVWR/NNCV	5/3	0.8/1.3	0.75/1.02

Figure 4.9 to 4.16 plot the relationship between w, m, matching with or without replacement, and RB of ACIE in 31 prevalence of treatments. In each figure, the panels on the left side are in response to the NNCVWR, and the panels on the right side are in response to the NNCV. Each two panels in the same row are in response to the same prevalence of treatments. Although we determine the optimal caliper width by the MSE of ACIE estimates, analysis of RB could help understand the variance and bias trade-off. In most of the cases, RB rapidly decreased to a minimum value as w increased from 0.15 to a point and then RB increased as m increased. If we compared Figures of RB with Figures of MSE, clearly, the optimal m and w based on MSE and RB are not consistent. For the same case, the optimal m and w based on MSE are always higher compared to those based on RB. It is expected because a point with the lowest bias usually corresponds to high variance. Another phenomenon is when RB increased as w rose from the optimal point, the curve of MSE tended to stabilize. It implies, and variance of ACIE estimates decreased as w increased the from the optimal point.

Regarding NNCVWR, the optimal w that minimize RB fell into the range between 0.15 and 0.9 in 29 out of 31 cases, and equal to 1.65 and 1.95 for the rest 2 cases. When using NNCV, the optimal w that minimize RB fell into the range between 0.15 and 0.95 in 27 out of 31 cases, and is between to 1.05 and 1.6 for the rest 4 cases. The optimal m that minimize RB is between 1 to 5 regardless the choice between NNCVWR and NNCV. It implies the optimal m depends on the prevalence of treatments but do not linearly associated with RB. When the prevalence of treatments is 0.2,0.8; 0.3,0.7 and 0.4,0.6, the RB is approximately constantly closed to 0. It indicates the choice of w had a much lower impact on the RB of ACIE estimates.

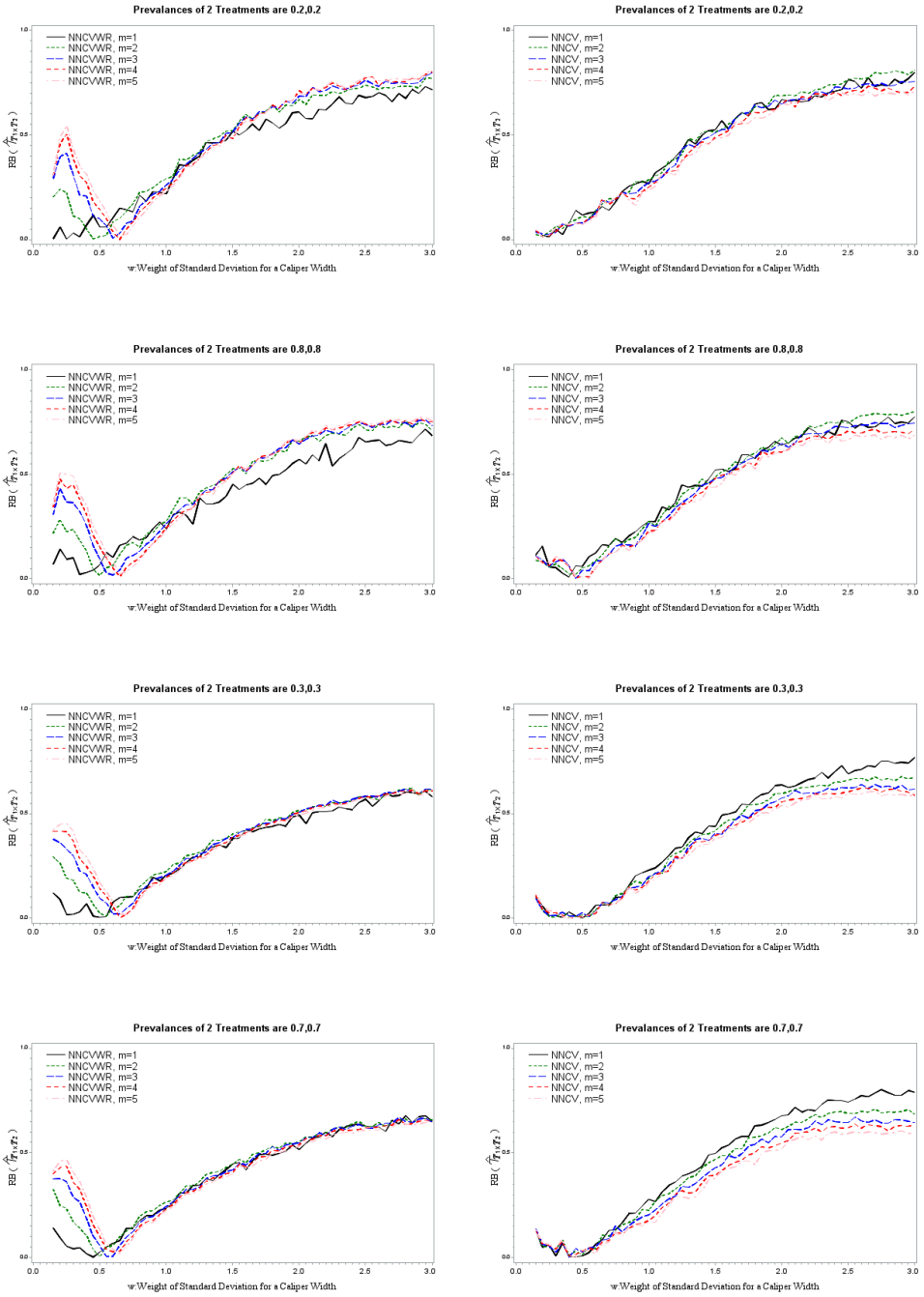

Figure 4.9 Caliper width, m vs. RB of ACIE Part 1

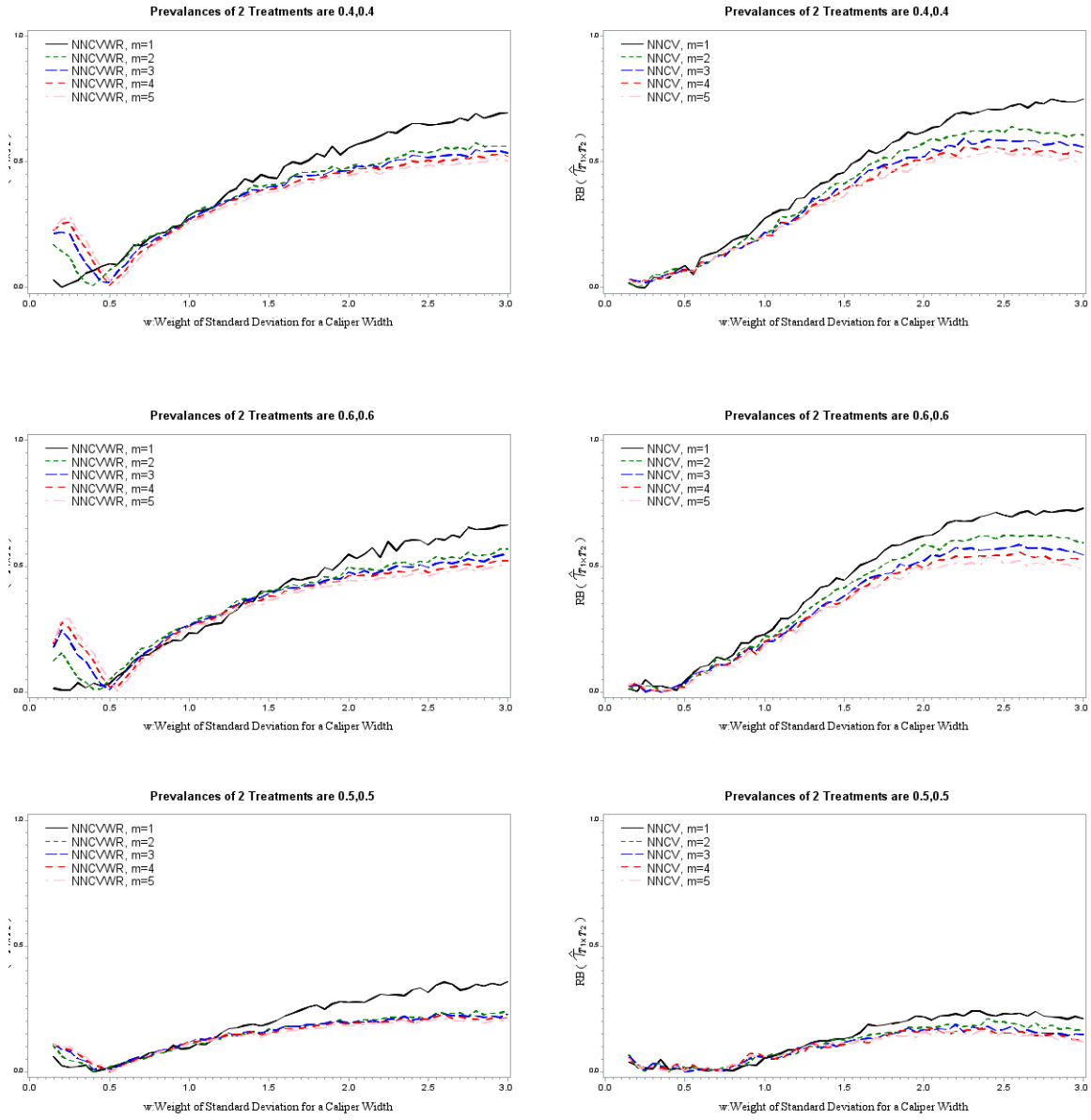

Figure 4.10 Caliper width, *m* vs. RB of ACIE Part 2

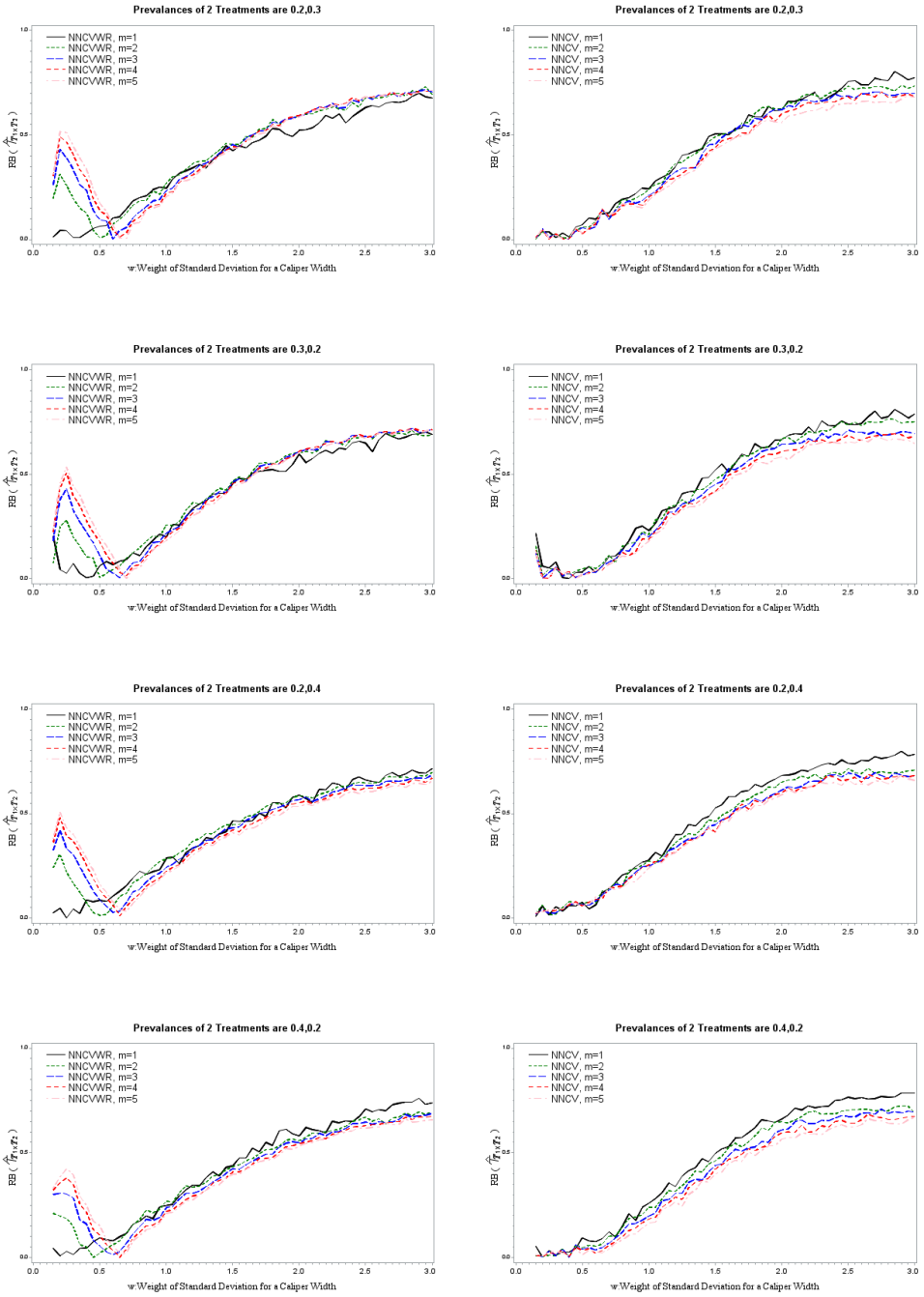

Figure 4.11 Caliper width, m vs. RB of ACIE Part 3

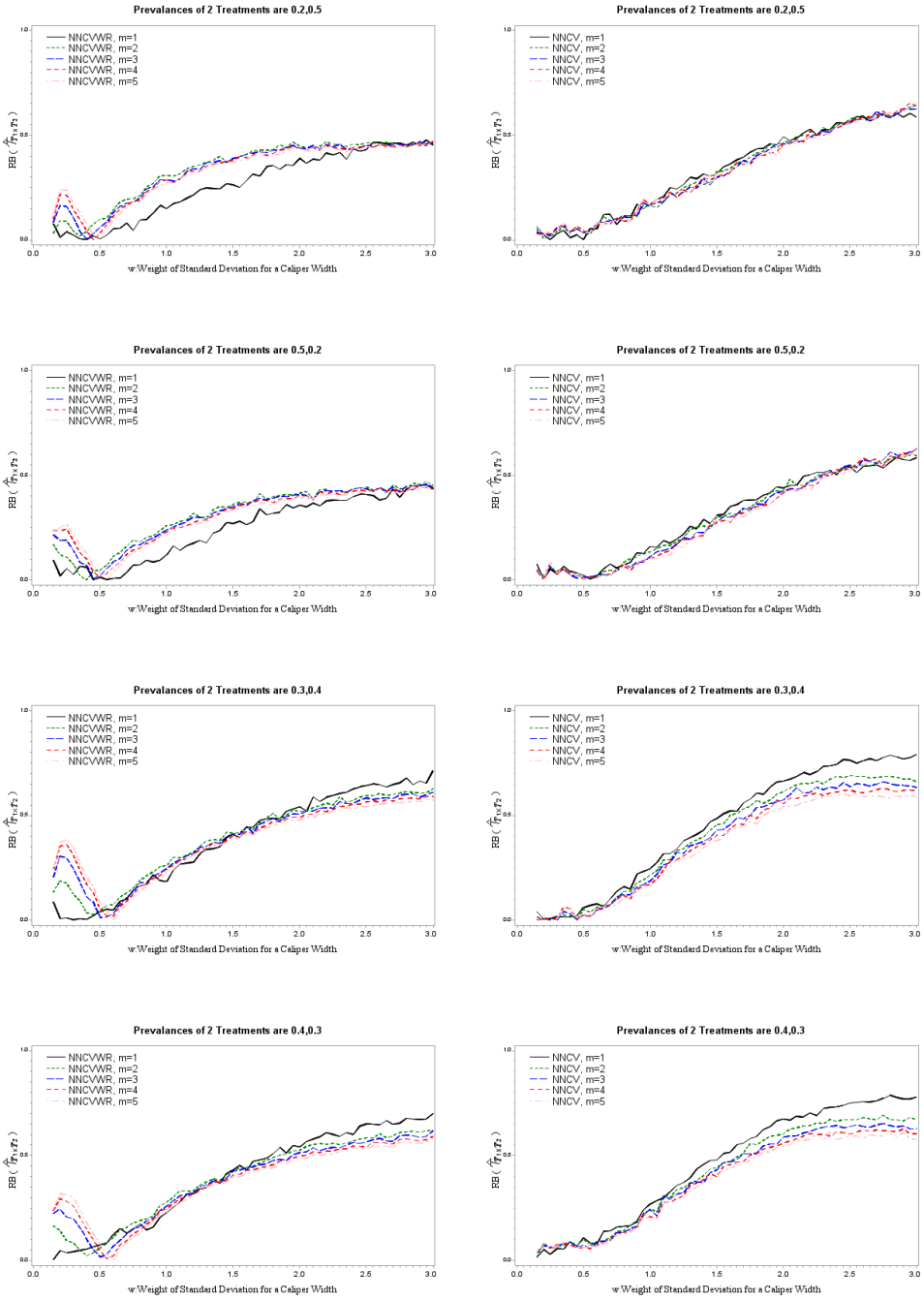

Figure 4.12 Caliper width, m vs. RB of ACIE Part 4

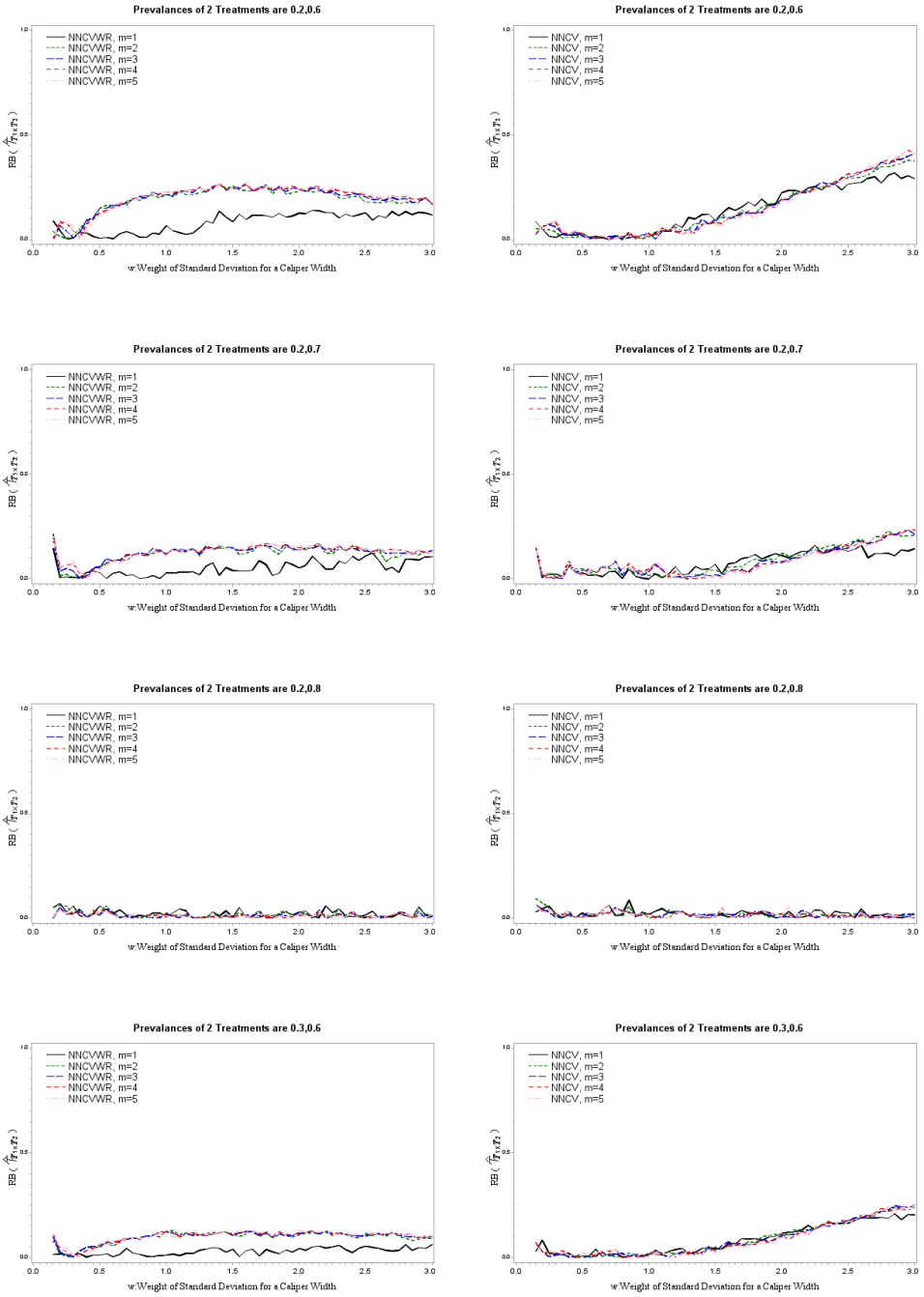

Figure 4.13 Caliper width, m vs. RB of ACIE Part 5

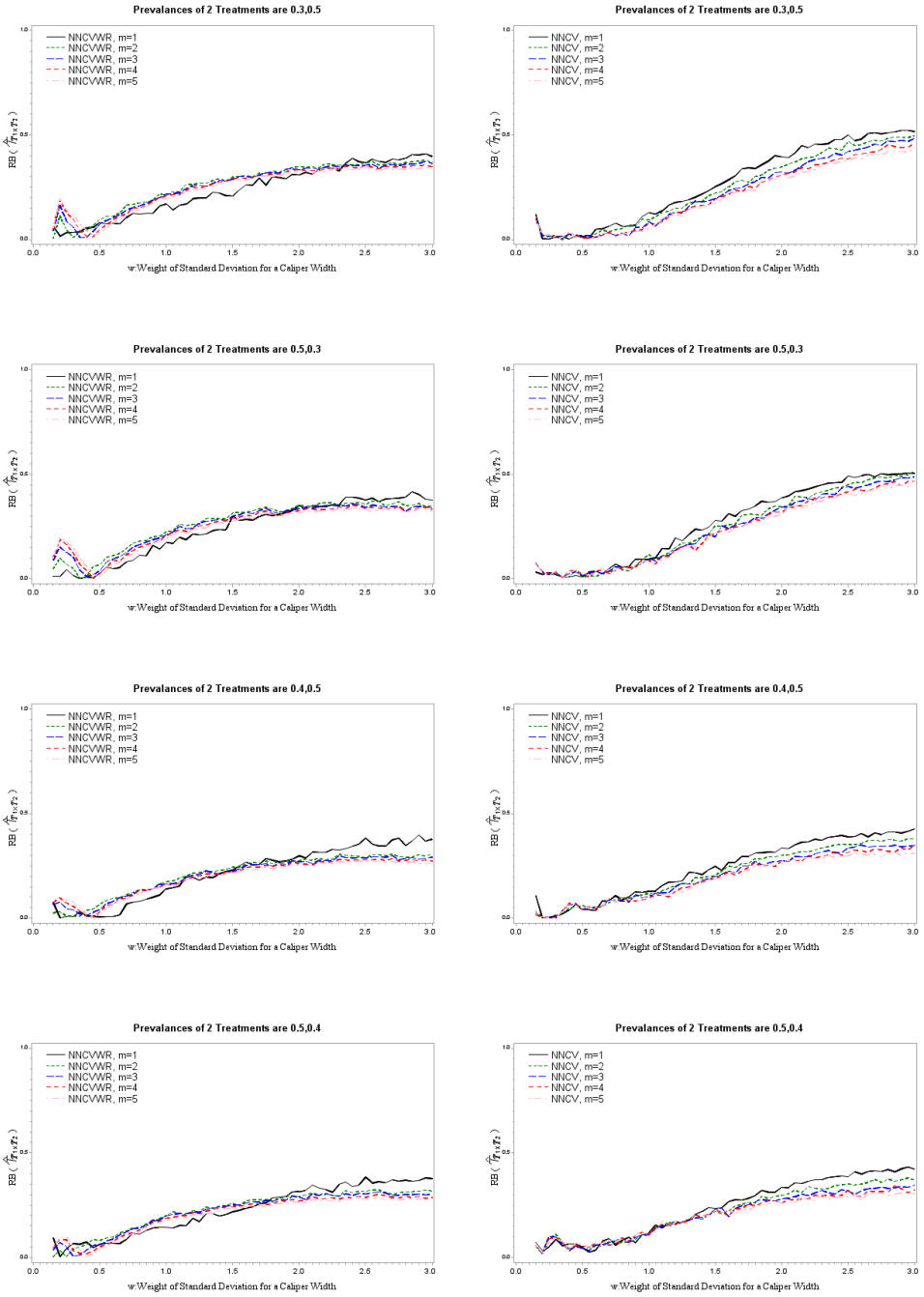

Figure 4.14 Caliper width, *m* vs. RB of ACIE Part 6

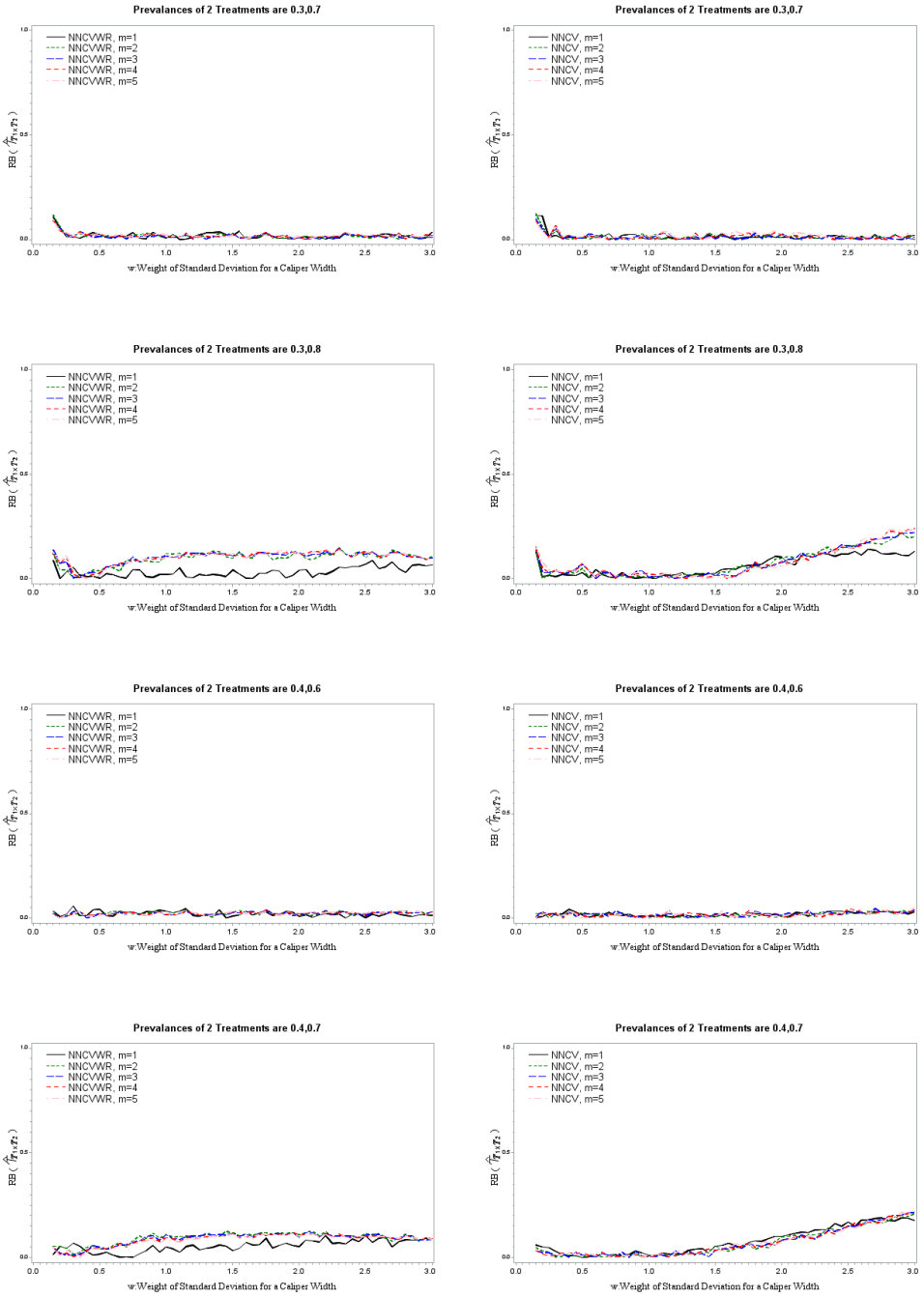

Figure 4.16 Caliper width, m vs. RB of ACIE Part 7

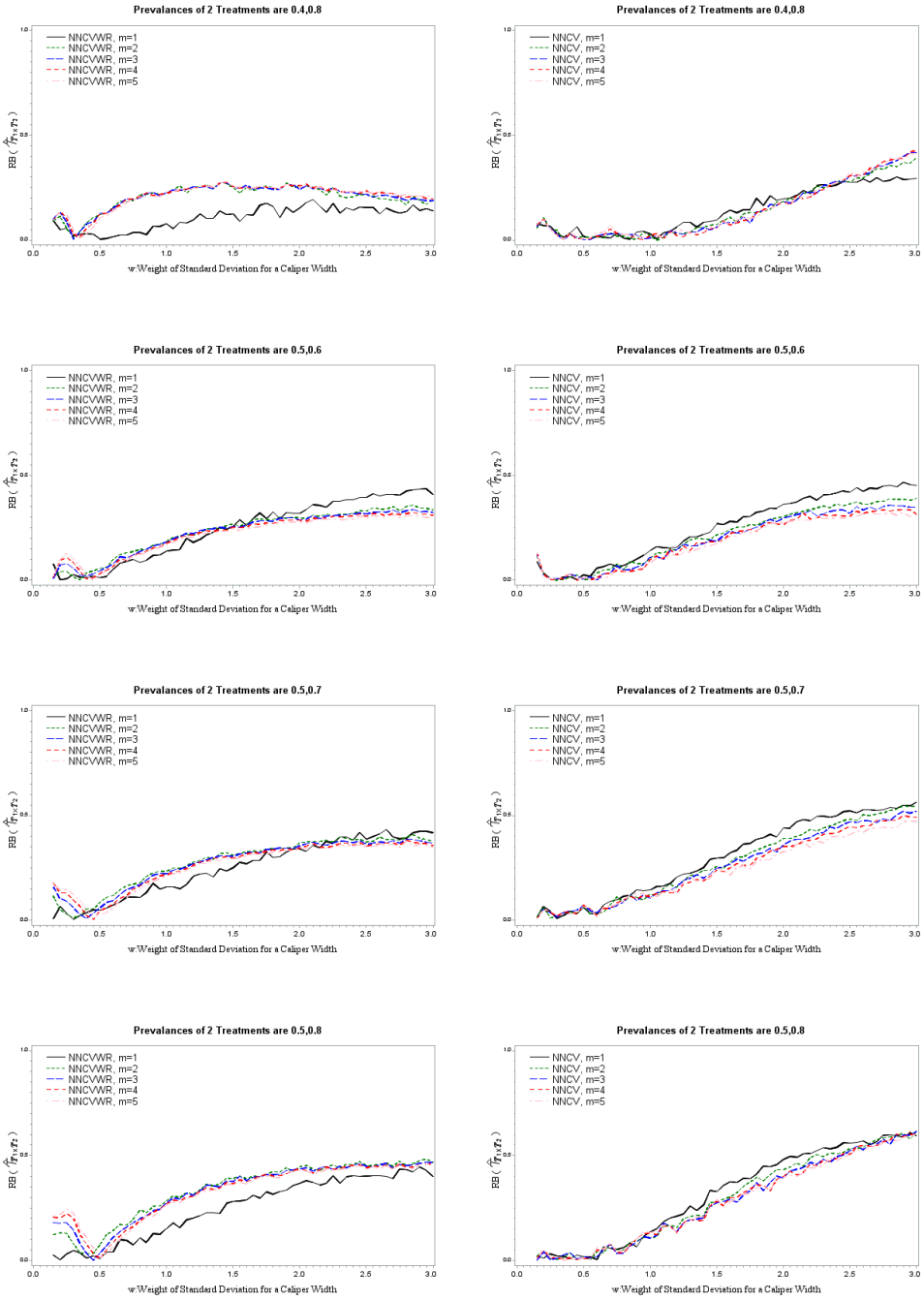

Figure 4.16 Caliper width, m vs. RB of ACIE Part 8

4.3.3 Simulation Results Part II

In this section, we compare the optimal NNCVWR and NNCV with the NN, NNWR, GIPTW, and 3x3 SP using linear regression with covariate adjustment discussed in section 3.4. We select the optimal NNCVWR and NNCV with the m and w minimizing the MSE of ACIE estimates in each prevalence of treatments scenario (see Table 4.9). We use MSE and RB of ACIE and ACME to evaluate these methods. Figure 4.17 to Figure 4.21 presents the comparisons of these six methods given 31 prevalence of treatment scenarios. The horizontal axis of each panel represents the prevalence of two treatments, and the vertical axis of each panel represents either MSE or RB of ACIE or ACME.

Cases: Prevalence of T_1 and T_2 are equal (Figure 4.17)

On the MSE of ACIE estimates, the NNCVWR tended to result in lower MSE than NN, NNWR, GIPTW, and 3x3 SP. The NNCVWR outperformed NNCV in 6 scenarios and was marginally weaker than NNCV in the cases of 0.5,0.5. The NNCV was superior to GIPTW, 3x3 SP, and NNWR, but inferior to NN in 4 cases. On RB of ACIE estimates, NNCVWR tended to result in RB less than 4% in 6 cases, and 8% in 1 case. In the case of 0.2,0.2, NNCVWR produced lowest RB, and NNCV produced lower RB than GIPTW, NNWR, and 3x3 SP. In the case of 0.3,0.3, NNCV tended to result in lowest RB, and NNCVWR was better than NN and 3x3 SP. In the case of 0.4,0.4, NNCV and NNCVWR had the nearly the same performance and outperformed GIPTW. In the case of 0.5,0.5, both GIPTW and NNCV had lower RB than other methods, while NNCVWR was marginally superior to NN. In the case of 0.6,0.6, NNCVWR had smallest RB as 3x3 SP,

and NNCV was better than NNWR. In the cases of 0.7,0.7, and 0.8,0.8, NNCVWR had lower RB than 3x3 SP, NNCV, and NNWR. When assessing MSE of ACME estimates, NNCVWR tended to result in smallest MSE in the cases of 0.4,0.4; 0.6,0.6; 0.7,0.7; and 0.8,0.8. NNCV was marginally inferior to NN. In the case of 0.2,0.2 and 0.3,0.3, NNCVWR had lower MSE than other methods except for GIPTW. In the case of 0.5,0.5, NNCV had the best performance, and NNCVWR was better than other methods except for NNCV. When assessing RB of ACME estimates, NNCVWR tended to yield approximately unbiased estimate in the cases of 0.2,0.2; 0.6,0.6; and 0.7,0.7, and RB less than 3% in other cases. NNCV tended to yield approximately unbiased estimate in the cases of 0.3,0.3; 0.5,0.5; 0.6,0.6; and 0.7,0.7, and RB less than 3.5% in other cases.

Cases: Prevalence of T_1 is 0.2 and unequaled to prevalence of (Figure 4.18)

In regards to MSE of ACIE estimates, NNCVWR and NN were similar and better than other methods in all cases expect 0.2,0.8. NNCV had lower MSE than GIPTW, 3x3 SP and NNWR in all scenarios, and close performance to NNCVWR in the case of 0.2,0.8. In regards to RB of ACIE estimates, NNCVWR had lowest RB in the case of 0.2,0.4, but higher RB than NNCV and NN in the cases of 0.2,0.6; 0.2,0.7; and 0.2,0.8. NNCV had higher RB than GIPTW and NN when the prevalence of T_2 was between 0.3 to 0.6. When assessing the MSE of ACME estimates for T_1, NN and NNCVWR were better than other methods and had almost the same performance except for the case of 0.2,0.8. NNCV was better than 3x3 SP, NNWR, and GIPTW in all cases except 0.2,0.3, where the MSE due to NNCV, 3x3 SP, and GIPTW were extremely closed from one to another. When assessing

the MSE of ACME estimates for T_2, NNCVWR, NN, GIPTW, and 3x3 SP had almost the same performance in all cases. NNCV was marginally weaker than NNCVWR, NN, GIPTW, and 3x3 SP. When assessing RB of ACME estimates, both NNCV and NNCVWR tended to yield approximately unbiased ACME estimates across all scenarios.

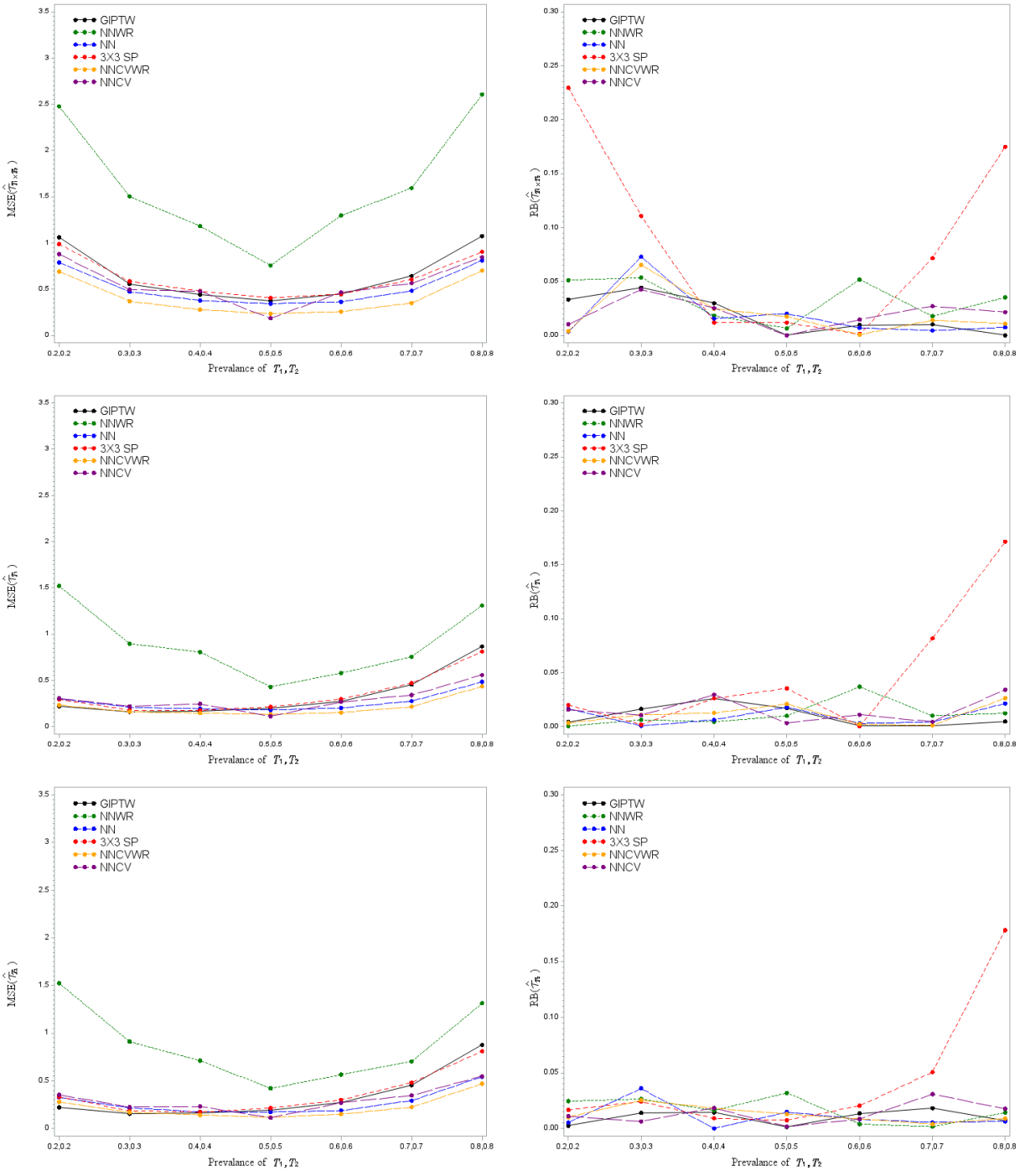

Figure 4.17 Case: Prevalence of T_1 and T_2 are equal

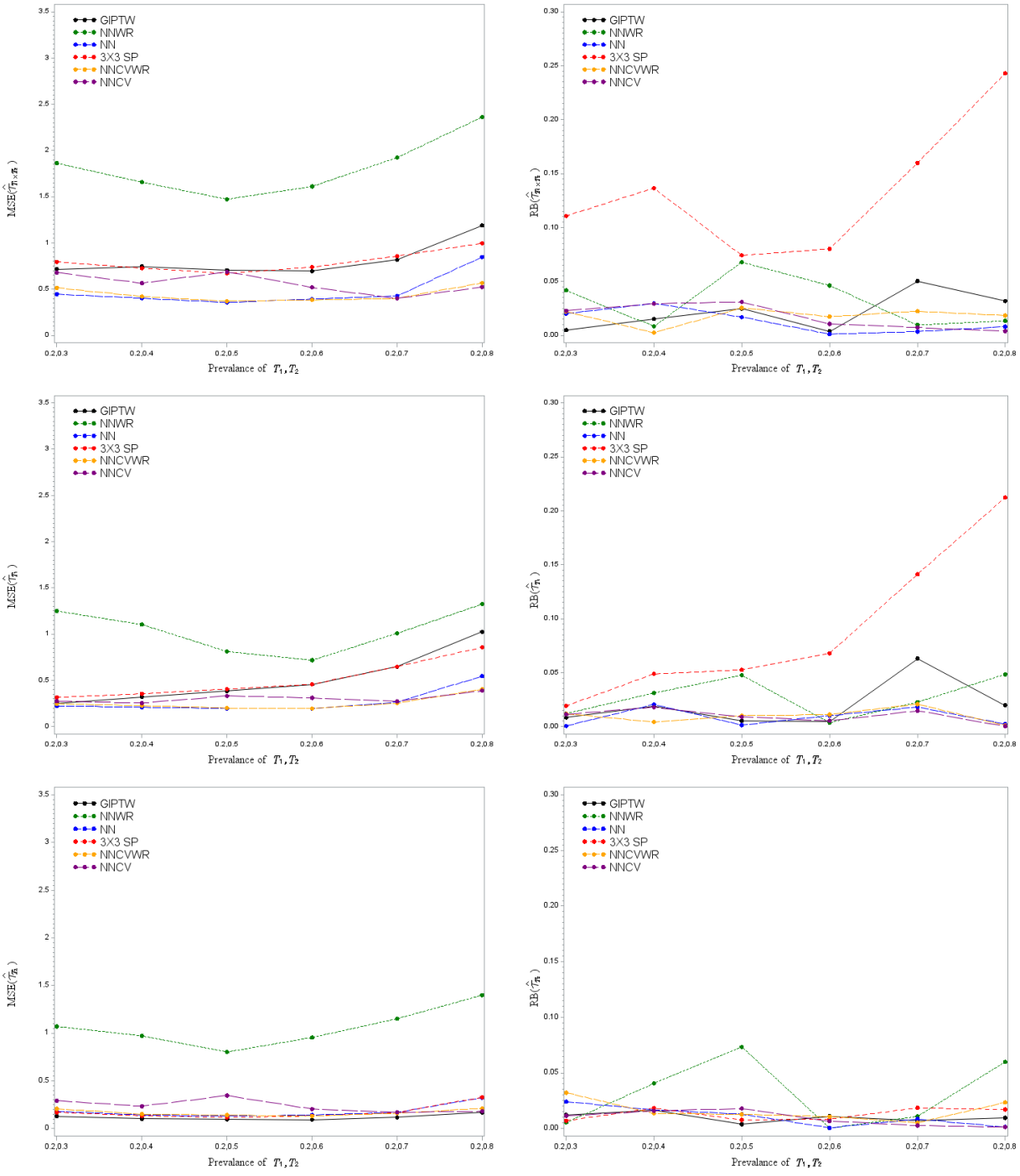

Figure 4.18 Case: Prevalence of T_1 equals 0.2 and unequal to Prevalence of T_2

Cases: Prevalence of T_1 is 0.3 and unequal to prevalence of T_2 (Figure 4.19)

When evaluating MSE of ACIE estimates, NNCVWR and NNCV were comparable and superior to other methods across in the cases of 0.3,0.6; 0.3,0.7; 0.3,0.8. NNCVWR had slightly higher MSE than NN in the cases of 0.3,0.2 and 0.3,0.4, and the same performance as NN in the case of 0.3,0.5. NNCV outperformed 3x3 SP, NNWR, and GIPTW in all cases. When evaluating RB of ACIE estimates, NNCVWR was inferior NNCV, NN, NNWR, and GIPTW in the case of 0.3,0.2. In the case of 0.3,0.4, NNCVWR had the same RB as NNCV and lower RB than 3x3 SP and NNWR. In the case of 0.3,0.5, NNCVWR had smaller RB than other methods except for NN. NNCV had the same performance as 3x3 SP and better performance than GIPTW and NNWR. In the case of 0.3,0.6, NN, NNCV and NNCVWR tended to result in smaller RB than other methods. In the case of 0.3,0.7, NNCVWR was the 2nd best approach, while NNCV was better than NNWR and 3x3 SP. In the case of 0.3,0.8, NNWR, NNCV, and NNCVWR coincided to the same point, and were better than 3x3 SP and GIPTW. When evaluating RB of ACME estimates for T_1, NNCVWR and NN tended to result in the same MSE, which was only marginally higher than the one for GIPTW. In the case of 0.3,0.4, NNCVWR was marginally weaker than NN and marginally superior to NNCV, 3x3 SP, and GIPTW. In the case of 0.3,0.5 and 0.3,0.6, both NN and NNCVWR tended to result in lowest MSE, while NNCV yielded slightly higher MSE than NN and NNCVWR. In the case of 0.3,0.7, both NNCV and NNCVWR had the lowest MSE. In the case of 0.3,0.8, NN, NNCV, and NNCVWR generated lower MSE than other methods. When evaluating RB of ACME estimates for T_1, both NNCV and NNCVWR tended to yield RB less than 2%. NNCVWR produced approximately unbiased estimate when the prevalence of T_2 was between 0.5 and 0.8. NNCV resulted in approximately unbiased estimate when the prevalence of T_2 was equal to 0.5, 0.7 and 0.8. When evaluating RB of ACME estimates for T_2, both NNCV and NNCVWR tended to yield RB less than 5% in the case of 0.3,0.2, but inferior

to NN, NNWR, and GIPTW. NNCV and NNCVWR produced nearly unbiased estimate when the prevalence of T_2 was equal to 0.4, 0.7 and 0.8. NNCV was slightly worse than other methods except for NNWR in the case of 0.3,0.5 and weaker than NN, NNCVWR, in the case of 0.3,0.6 and 0.3,0.7.

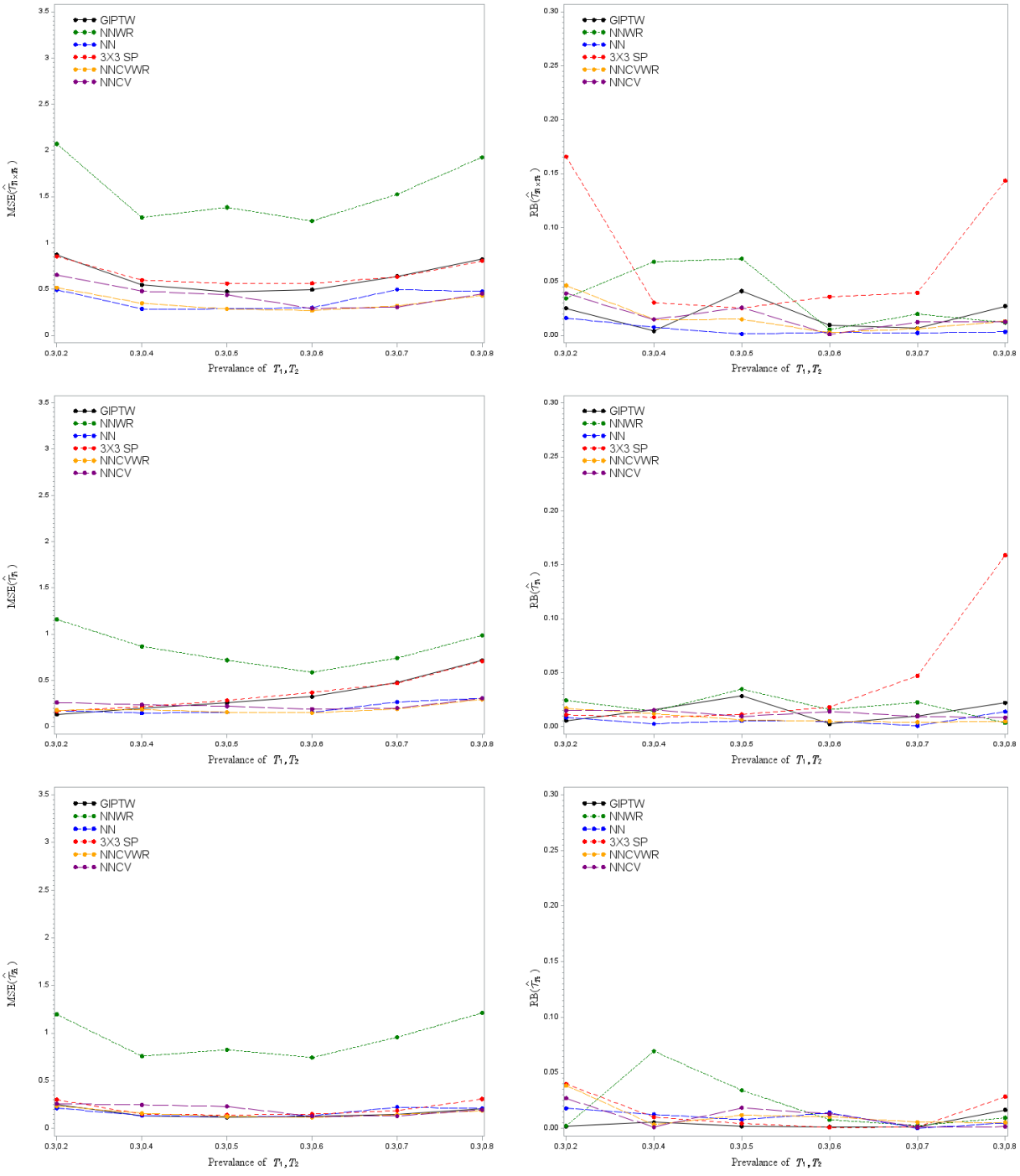

Figure 4.19 Case: Prevalence of T_1 equals 0.3 and not equal to Prevalence of T_2

Cases: Prevalence of T_1 is 0.4 and not equal to prevalence of T_2 (Figure 4.20)

When examining MSE of ACIE estimates, NNCVWR had similar MSE to NN and outperformed other approaches across all cases. NNCV had the 3rd lowest MSE when the prevalence of T_2 was equal to 0.2, 0.3, 0.5, 0.7 and 0.8. NNCV was superior to other methods except for NNCVWR in the case of 0.4,0.6. When examining RB of ACIE estimates, both NNCVWR and NNCV tended to result in RB less than 3% in all cases. In the case of 0.4,0.2, NNCVWR had lower RB than NNCV, NNWR, and 3x3 SP, while NNCV was better than NNWR and 3x3 SP. In the case of 0.4,0.3, NNCV tended to produce the lowest RB, while NNCVWR was marginally better than NNWR and NN, and much better than 3x3 SP. In the case of 0.4,0.5, both 3x3 SP and NNCVWR had lowest RB, while NNCV was only better than NNWR. In the case of 0.4,0.6, NNCV yielded lower RB other methods, while NNCVWR was only slightly inferior to NNCV. In the case of 0.4,0.7, both NNCV and NNCVWR were weaker than NNWR and GIPTW. In the case of 0.4,0.8, GIPTW, NN, NNCVWR, and NNCV had nearly the same RB, and lower RB than 3x3 SP, and NNWR. When examining MSE of ACME estimates, NNCVWR tended to result in lowest MSE is all cases except 0.4,0.2. In the case of 0.4, 0.2, NNCVWR had slightly lower MSE than 3x3 SP and GIPTW. NN, NNCV, and NNCVWR were pretty comparable across all cases. When examining RB of ACME estimates, both NNCVWR and NNCV tended to result in RB less than 2% in all cases.

Cases: Prevalence of T_1 is 0.5 and not equal to prevalence of T_2 (Figure 4.21)

When assessing MSE of ACIE estimates, NNCVWR and NN were similar with lower MSE than other methods. NNCV had the 3rd lowest MSE among six methods. In regards to RB of ACIE estimates, NNCV had lowest RB in the cases of 0.5,0.2 and 0.5,0.7. NN had lowest RB in the case of 0.5,0.8, and second lowest RB in the case of 0.5,0.3. When assessing MSE of ACME estimates, NNCVWR and NN had nearly the same performance and outperformed other methods in most of the cases. NNCV had smaller MSE than

GIPTW, NNWR, and 3x3 SP when the prevalence of T_2 is between 0.6 to 0.8. In regards to RB of ACME estimates for T_1, NNCVWR tended to yield approximately unbiased estimate when the prevalence of T_2 was equal to 0.2, 0.3, 0.6 and 0.7. NNCVWR was inferior to other methods except for GIPTW in the case of 0.5,0.4. NNCV tended to yield approximately unbiased ACME estimate in the cases of 0.5,0.7 and 0.5,0.8. In regards to RB of ACME estimates for T_2, both NNCV and NNCVWR tended to result in RB less than 2%. NNCV had lowest MSE when the prevalence of T_2 was equal to 0.2, 0.4, 0.6, and 0.7. NNCVWR had very closed RB to NNCV and outperformed other methods in the case of 0.5,0.3.

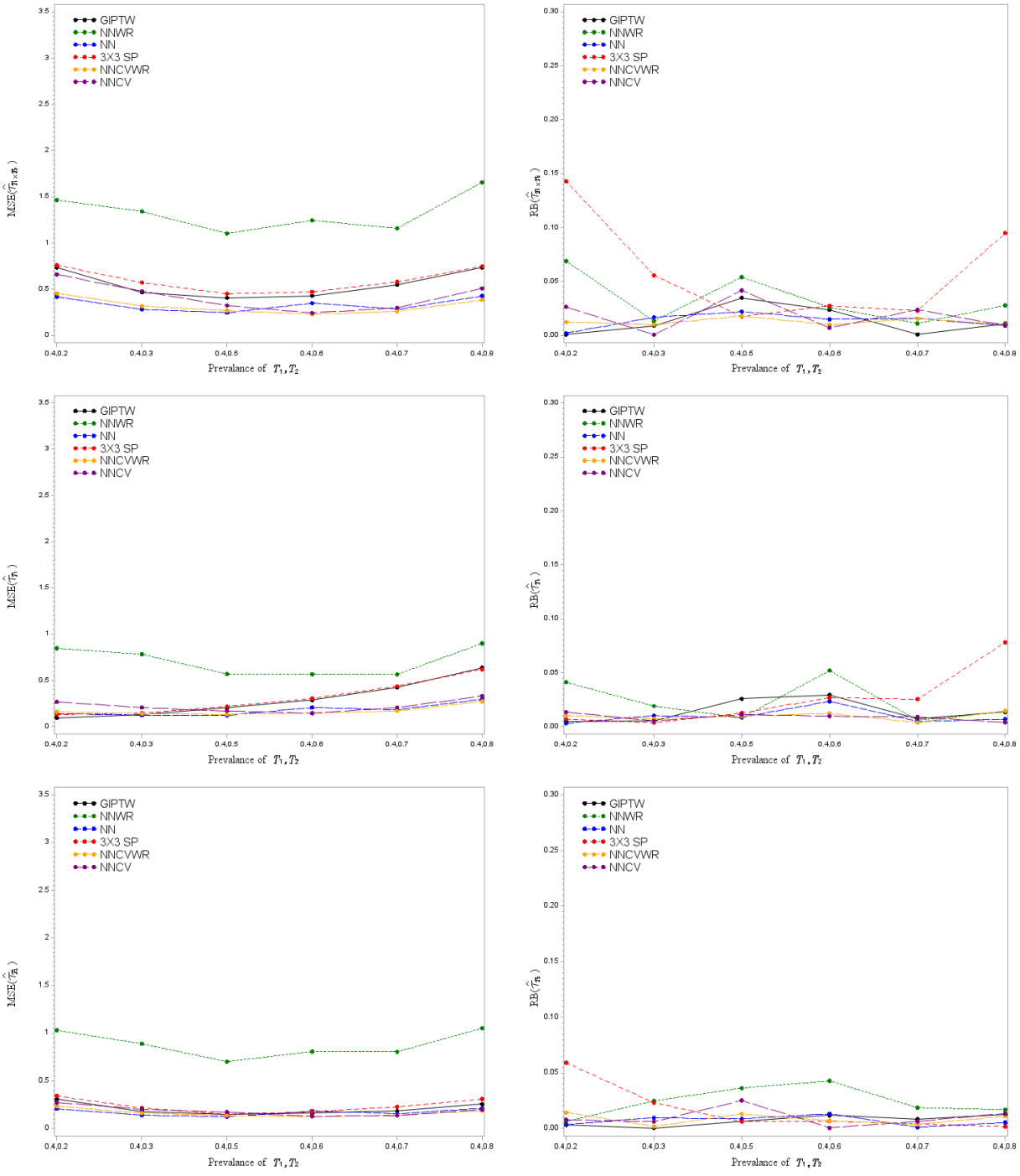

Figure 4.20 Case: Prevalence of T_1 equals 0.4 and not equal to Prevalence of T_2

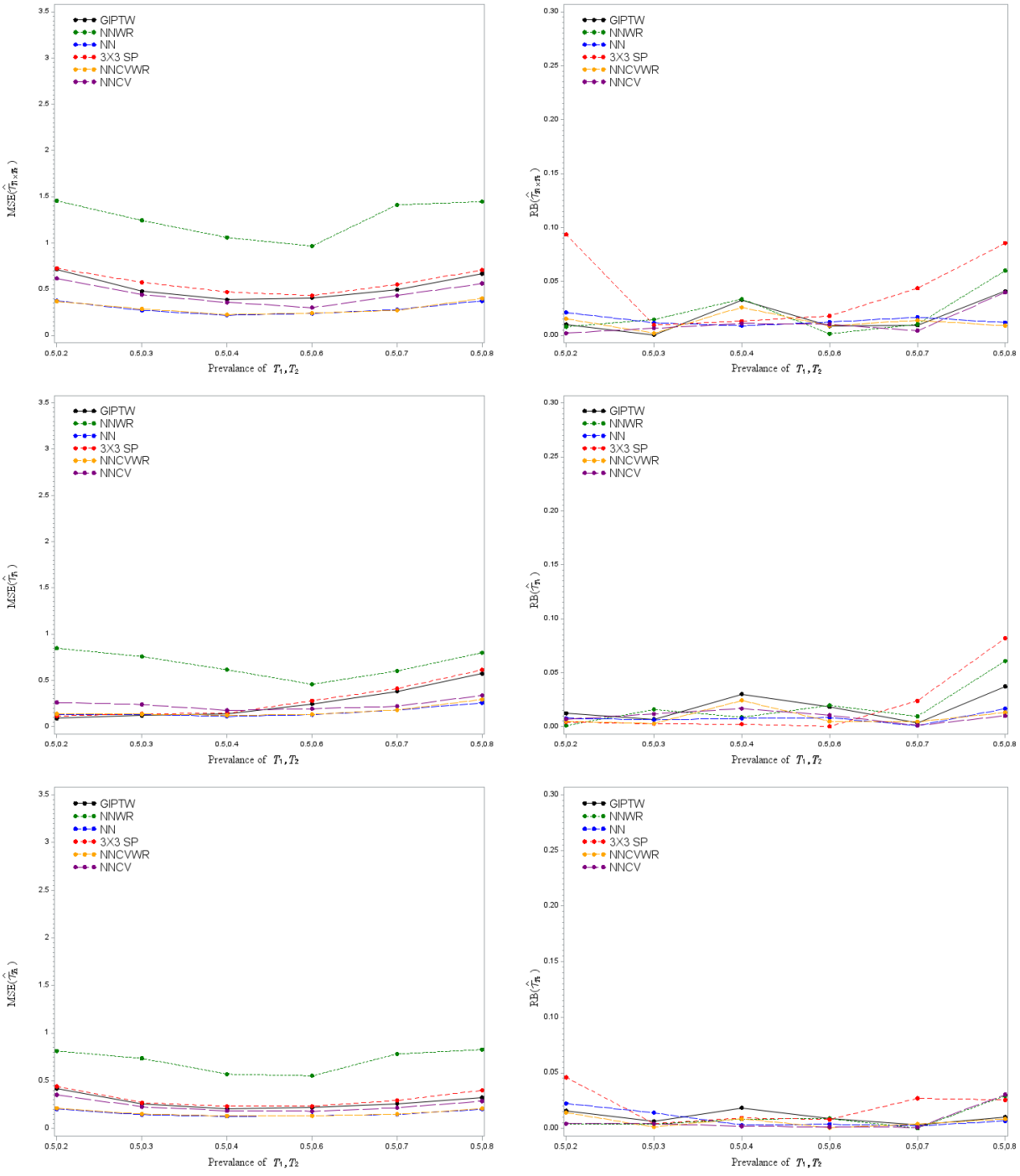

Figure 4.21 Case: Prevalence of T_1 equals 0.5 and not equal to Prevalence of T_2

4.5 Case Study: Effect of Smoking on National Medical Expenditure Using a Bivariate Treatment

4.5.1 Introduction

In previous simulation studies, we showed NN and NNCVWR could potentially improve ACIE and ACME estimation. In this section, we turn to real observational data. we employ NNCVWR, NN, GIPTW and 3x3 SP to the 1987 National Medical Expenditure Survey (NMES) data. We estimate the average causal main effects (ACME) of frequency and duration of smoking as well as their average causal interaction effect (ACIE) on medical expense. There are a number of literatures using propensity score methods (PSM) to estimate the causal effects of smoking on medical expenditure (e.g., Larsen 1999; Zeger, Wyant, Miller, and Samet 2000). These papers, however, merely concentrated on the difference between smokers and nonsmokers on medical cost but did not consider other forms of treatment such as duration and frequency of smoking. Motivated by this restriction, Imai and van Dyk (2004) proposed P-Function theory and stratification on P-Function (SP). They provided two examples of estimating the effect of smoking using SP. In the first example, they used a continuous treatment variable, namely packyear defined by Johnson et al. (2003) as $\log(packyear) = \log(\frac{frequency}{20 \cdot duration})$, where frequency represents the number of cigarettes smoked per day, and duration represents the reported number of years smoked. They estimated the P-Function of $\log(packyear)$ using a linear regression model with covariates adjustment. Then they equally stratified on the estimated P-Function using 3, 5 and 10 strata respectively. Within each subclass they fitted

linear regression to estimate the within-subclass average causal effect (ACE) of $\log(packyear)$ on medical expense. They then computed the weighted average of the within-subclass ACE estimates to obtain the overall ACE. They found that smoking tended to significantly increase medical cost. In the second example, they decomposed the $\log(packyear)$ into two continuous treatment variables, which are duration and frequency of smoking. They estimated the P-Functions for duration and frequency of smoking using linear regression models. Instead of stratifying on the $\log(packyear)$, they separately and equally stratified the estimated P-Functions for duration and frequency of smoking by 3 or 4 subclasses, namely, 3x3 SP or 4x4 SP. They regressed the outcome on the two treatment variables controlling for covariates using linear regression to estimate the within-subclass ACE. The overall ACE estimate was then calculated by the weighted average of within-subclass ACE. Both 3x3 SP and 4x4 SP indicated only frequency of smoking significantly increased medical expense. They also showed in simulations that SP reduced bias of ACE estimates as compared to directly using linear regression models. Despite the appealing theoretical properties of SP, they did not consider the potential interaction effect between duration and frequency of smoking. Therefore, we aim to estimate both ACIE and ACME using NN, NNCVWR, 3x3 SP, and GIPTW and compare their results.

4.5.2 Data Description

In 1987, the US Department of Health and Human Services, Public Health Service conducted National Medical Expenditure Survey on annual medical expenditures and disease status for a representative sample of the U.S. civilian, non-institutionalized

population. For comparison purpose, we use the data cleaned by Imai and van Dyk (2004). In particular, an individual was excluded if: 1) the individual was a non-smoker; 2) the individual had missing information; 3) the individual excluded non-smokers and smokers with any missing values, and they also The data includes 9073 smokers after excluding non-smokers and smokers with any missing values. The primary outcome is medical expenditure. The covariates of interests are age at the survey, age when smoking began, sex (female, male), race (black, white, other), marital status (single, separated, divorced, widowed, married), poverty status (high income, middle income, low income, near poor, poor), education level (high school, some college, bachelor), census region (West, South, Midwest, Northeast), and seat belt usage (always, sometimes, rarely). There are two continuous exposure variables: frequency and duration of smoking.

4.5.3 Data Analysis

In this study, we dichotomize frequency and duration of smoking into two binary exposure variables T_1: whether the amount of cigarettes smoked each day is greater than 19 and T_2: whether the years of smoking is greater than 25. We determine the optimal cut-off points of two continuous treatment variables by searching for the maximum value of Kendall rank correlation coefficient (Kendall, 1938) across all possible cut-offs over the space of original continuous treatment variables. We change the form of treatments because of two reasons: First, our matching algorithm is confined to continuous treatment variables. Second, although the NNCVWR can be applied to multiple ordinal or categorical treatments variables, the optimal NNCVWR identified from simulation results in section 4.3 was based on two binary treatments. After dichotomizing two treatment variables, the

prevalence of T_1 is 0.543, and the prevalence of T_2 is 0.417. Since the prevalence of two treatments is closed to 0.5,0.4, we use the optimal NNCVWR with $w = 0.9, m = 5$ (see table 4.9) for this study.

We estimate the P-Function using a bivariate logistic regression model including the covariates described in section 4.5.2. In addition, we included the squared terms of age at the survey and age when smoking started. For comparison purpose, the model specification is the same as what Imai and van Dyk (2004) did. When estimating the generalized propensity score (GPS) for GIPTW, we first combine two treatment variables into a categorical variable with 4 levels, then we estimated the generalized propensity score as the probability of falling into the actual exposure group. After propensity score estimation, we apply each method specified in previous sections (see section 3.2.2 for NN, section 3.3.1 for 3x3 SP, section 3.3.2 for GIPTW, section 4.2.3 for NNCVWR). After using NN, we estimate the ACIE and ACME using linear regression on the matched data. We regress the outcome on a constant, T_1, T_2, the interaction term between T_1 and T_2, and all of the covariates. The model specifications of other methods are consistent with NN. After using NNCVWR, we apply weighted linear regression to the matched data. The computation of weight is referred to section 4.2.3. For GIPTW, we use the inverse of GPS as the weight, and apply weight linear regression to the original data. For 3x3 SP, we follow Imai and van Dyk (2004) who separately and equally divide the distribution of propensity score estimates of T_1 and T_2 into 3 subclasses. We then estimate then ACIE and ACME within each strata using linear regression. We calculate the overage ACIE and ACME by taking

the weighted average of within-subclass estimates. The standard error of overall ACIE or ACME is equal to the weighted average of within-subclass standard errors. The statistical tests of whether the ACIE or ACME is different from 0 is under 0.05 significance level. for all statistical analysis was implemented using SAS 9.3.

4.5.4 Results

Table 4.10 reports the results of 4 methods. All methods indicate there is no significant interaction between duration and frequency of smoking. NN, NNCVWR, and GIPTW found cigarettes smoked greater than 19 per day significantly increased medical cost. Although the binary treatments variables are less informative as compared to the continuous treatment variables, the results of both NN and NNCVWR agree to Imai and Imai and van Dyk (2004)'s results estimated using continuous treatments. On the other hand, by changing continuous variables to binary variables, the 3x3 SP did not find any significance. It implies the balance due to 3x3 SP could be inferior to other methods, which is also consistent to our simulation results.

Table 4.10 Summary of ACME and ACIE Estimates (Standard Errors) of Increased Smoking on Medical Expenditure by Methods

Method	T_1	T_2	$T_1 \times T_2$
3x3 SP	0.102(0.325)	0.295(0.284)	-0.206(0.463)
GIPTW	0.113(0.053)*	0.156(0.058)*	-0.076(0.073)
NN	0.051(0.065)	0.205(0.067)*	-0.094(0.089)
NNCVWR, m=5, w=0.9	0.125(0.087)	0.244(0.071)*	-0.183(0.114)

Note: T_1 is an indicator of frequency of smoking ≥ 20, T_2 is an indicator of duration of smoking ≥ 26 years, and their standard errors (In brackets) are reported. * indicates p-value<0.05.

4.6 Discussion

In this chapter, I introduced two generalized propensity score methods: NNCV and NNCVWR stimulated by the bipartite nearest neighbor caliper width matching (Cochran and Rubin, 1974; Austin, 2011) and variable matching (Ming and Rosenbaum, 2000). Both of them apply to artificial treatment scheme except continuous treatments. The balance after matching between treatment groups due to NNCV or NNCVWR is controlled by three factors: 1. a weight of standard deviation of all possible pairwise Euclidian distances computed using P-Function estimate; 2. A threshold m that controls that a maximum number of units can be selected from a treatment group for a matched set; 3 Matching with or without replacement. I used a series of Monte Carlo simulations to find out the optimal weight, and m for NNCV and NNCVWR in the various prevalence of treatment settings given two binary treatments. We determined the optimal caliper width by the minimum MSE of ACIE estimates. When using NNCVWR, the optimal m is always to 5 given $1 <= m <= 5$, and the optimal weight is ranged from 0.6 to 1 in 71% cases, and from 1.9 to 3 in 29% cases. When using NNCV, the optimal m is 3 in 58% cases, 2 in 32% cases, and 3 in 10% cases given $1 <= m <= 5$. The optimal w is ranged from 0.9 to 3. The optimal NNCVWR tended to have superior performance than optimal NNCV in 93.75% cases. We

further compared optimal NNCV and optimal NNCVWR to NN, NNWR, 3x3 SP, and GIPTW using simulations. When estimating MSE of ACIE, the optimal NNCVWR tended to result in estimates with the greatest precision in 16 out of 31 cases, optimal NNCV in 3 cases, and NN in 12 cases. When estimating MSE of ACME for T_1, optimal NNCVWR tended to yield estimate with higher precision in 10 cases, optimal NNCV in 3 cases, NN in 12 cases, and GIPTW in 6 cases. When estimating MSE of ACME for T_2, optimal NNCVWR was superior to other methods in 9 cases, optimal NNCV in 3 cases, NN in 10 cases, and GIPTW in 9 cases. We found a variable matching with replacement could improve MSE and RB of ACIE and ACME estimates in a few scenarios compared to NN. When estimating RB of ACIE estimates, NNCVWR showed superior performance in 3 cases; NNCV led in 8 cases, and NN in 11 cases. When estimating RB of ACME estimates for T_1, NNCVWR showed the best performance in 2 cases, NNCV led in 6 cases, NN in 8 cases, 3x3 SP was superior to other methods in 5 cases, and GIPTW led in 4 cases. In applied data analysis, NN and NNCVWR resulted in consistent results with previous literature.

In practice, given the data is similar to our simulation settings and the ACIE is of primary interests, then we suggest the optimal NNCVWR or NN. The choice between them depends on the prevalence of treatments. If the ACME is of primary interests, we recommend choosing one from NNCVWR, NN, NNCV or GIPTW according to the prevalence of treatments. We encourage researchers to refer to Table 4.11 to decide the optimal solution.

There are certain limitations to our study. First, the choice of m was set to less than 6. It is possible that NNCVWR would be improved with m greater than 5. Second, our simulation settings did not consider binary and categorical outcomes and covariates. Third, although NNCVWR and NNCV apply to multivariate categorical and ordinal treatments, our simulation settings only include the simplest extension: 2 binary treatments. In future, it is desirable to examine these methods given other types of treatment variables.

Table 4.11 Recommendations on Choice of Matching Algorithm based on Different Performance Measures by Prevalence of T_1, T_2

Prevalence of T_1, T_2	**MSE of ACIE**	**MSE of ACME for T_1**	**MSE of ACME for T_2**
0.2,0.2	Optimal NNCVWR	GIPTW	GIPTW
0.2,0.3	NN	NN	GIPTW
0.2,0.4	NN	NN	GIPTW
0.2,0.5	NN	NN	GIPTW
0.2,0.6	Optimal NNCVWR	NN	GIPTW
0.2,0.7	Optimal NNCVWR	Optimal NNCVWR	GIPTW
0.2,0.8	Optimal NNCV	Optimal NNCV	GIPTW
0.3,0.2	NN	GIPTW	NN
0.3,0.3	Optimal NNCVWR	GIPTW	GIPTW
0.3,0.4	NN	NN	NN
0.3,0.5	Optimal NNCVWR	NN	GIPTW
0.3,0.6	Optimal NNCVWR	Optimal NNCVWR	Optimal NNCVWR
0.3,0.7	Optimal NNCV	Optimal NNCVWR	Optimal NNCV
0.3,0.8	Optimal NNCVWR	Optimal NNCVWR	Optimal NNCVWR
0.4,0.2	NN	GIPTW	NN
0.4,0.3	NN	NN	NN
0.4,0.4	Optimal	Optimal NNCVWR	Optimal NNCVWR

	NNCVWR		
0.4,0.5	NN	NN	NN
0.4,0.6	Optimal NNCVWR	Optimal NNCV	Optimal NNCVWR
0.4,0.7	Optimal NNCVWR	Optimal NNCVWR	Optimal NNCV
0.4,0.8	Optimal NNCVWR	Optimal NNCVWR	Optimal NNCVWR
0.5,0.2	Optimal NNCVWR	GIPTW	NN
0.5,0.3	NN	GIPTW	NN
0.5,0.4	NN	NN	NN
0.5,0.5	Optimal NNCV	Optimal NNCV	Optimal NNCV
0.5,0.6	NN	NN	Optimal NNCVWR
0.5,0.7	Optimal NNCVWR	NN	NN
0.5,0.8	NN	NN	NN
0.6,0.6	Optimal NNCVWR	Optimal NNCVWR	Optimal NNCVWR
0.7,0.7	Optimal NNCVWR	Optimal NNCVWR	Optimal NNCVWR
0.8,0.8	Optimal NNCVWR	Optimal NNCVWR	Optimal NNCVWR

Note: The 2nd column addresses the best approaches by the lowest MSE of ACIE estimates. The 3rd column addresses the best approaches by the minimum MSE of ACME estimates for T_1. The 4th column addresses the best approaches by the smallest MSE of ACIE estimates for T_2.

Chapter 5. Concluding Remarks and Future Work

5.1 Discussion

The main limitation of propensity score methods (PSM) of Rosenbaum and Rubin (1983) is it cannot be applied to non-binary treatments. In Chapter 1 we reviewed the history of PSM along with two existing generalizations of PSM: stratification on P-Function (SP) (Imai and van Dyk, 2004) and generalized inverse probability treatment weighting (GIPTW) (Imbens, 2000); We pointed out previous studies did not provide theorems and examinations of average causal interaction effect (ACIE), which sometimes could be more crucial than estimating the average causal main effect (ACME). In chapter 2 we established theoretical basis for generalized matching methods based on P-Function framework. In particular, we derived the formula for percent bias reduction given the P-Function. In addition, we showed under minor assumptions, the percent bias reduction on P-Function is proportional to the percent bias reduction for any dimension of the covariate distribution. Moreover, we provided theorems of estimating ACIE and ACME given two binary treatments. In chapter 3, we introduced a distance measure that could summarize the similarity structure among multiple treatment groups. Based on the distance measure we provided two generalized propensity score matching strategies: NN and NNWR. These two methods are extended from nearest neighbor matching with and without replacement and are applicable for any form of treatment except continuous treatments. In a simulation study, we showed that NN could potentially improve the statistical properties of estimated ACIE and ACME compared to GIPTW and 3x3 SP given two binary treatments, whereas

we do not suggest use NNWR it tended to result in greater MSE of ACIE and ACME estimates than those of other 3 methods in most of scenarios. In chapter 4, we further proposed another two generalization of propensity score matching algorithms: NNCVWR and NNCV. We demonstrated the quality of matching due to NNCV or NNCVWR is controlled by three factors: 1) w :a weight of standard deviation of all possible pairwise Euclidian distances computed using P-Function estimate; 2) A threshold m that controls that maximum number of units can be selected from a treatment group for a matched set; 3) Matching with or without replacement. To better utilize these two methods in practice, we conducted simulations to find out the optimized w and m for NNCVWR and NNCV in a variety of cases. The simulation results indicate NNCVWR outperformed other methods when evaluating the MSE of ACIE estimates in 51.6% cases and had pretty similar performance to the best solution in the rest of cases. Besides, NNCVWR was superior to other methods on MSE of ACME estimates in over 20% cases, and closed to the optimal solutions in other cases. With our approaches, researchers could estimate ACME and ACIE by matching on a low-dimensional parameterizations of the P-Function rather than on high-dimensional baseline covariates. In an applied example, we estimated the ACME of frequency and duration of smoking and their ACIE on national medical expenditure using the NN, NNCVWR, 3x3 SP, GIPTW along with the NEMS data. The results using both approaches indicated that only frequency of smoking significantly increased the medical expense. Although we lost information after dichotomizing two treatment variables, the results using NN and NNCVWR were consistent with Imai and van Dyk's

(2004) results. However, we also found their methods end up with different conclusion when the continuous treatments changed to binary treatments.

There are several limitations on this dissertation. First, the P-Function is always unknown in real data, and our simulation study only considered the situation when the model specification of P-Function estimation is correctly specified. Nevertheless, minor misspecifications of propensity score model could lead to substantial bias of estimated treatment effects (Kang and Schafer, 2007; Smith and Todd, 2005). Imai (2014) proposed covariate balancing propensity score (CBPS) that optimize the covariate balance when estimating the P-Function. He proved that CBPS could leverage the performance of propensity score matching methods. Therefore, we suggest researchers employ CBPS to estimate P-Function, and then apply our matching strategy. In the future, it is valuable to examine whether our methods could be improved when combing with CBPS. Second, in terms of program run time, the NN had the longest run time among all methods. NNCVWR is much faster than NN but slower than 3x3 SP and GIPTW. In the future, we could further advance their speeds using parallel computation. A third limitation is our approaches are under the frequentist paradigm to date. We did not consider modeling uncertainty in the P-Function. Researchers discussed Bayesian propensity score matching methods to solve this problem (e.g. Rubin, 1985; Lawrence et al, 2009; An, Weihua. 2010). In future research, we could further extend our matching methods to the Bayesian paradigm. Here I briefly introduce the idea. We could estimate the posterior distribution of P-Function using MCMC methods. For example, suppose the MCMC process includes 1000 iterations, for each iteration we generate a sample of estimated P-Functions. For each

sample we could apply NN or NNCVWR to obtain a matched data set, then estimate ACIE and ACME using regression with covariate adjustment. Therefore, we could obtain 1000 ACIE and ACME estimates, and calculate credible intervals and posterior mean of ACIE and ACME estimates. Finally, we implemented these methods using SAS macro. In the future, we will develop packages of NN and NNCVWR in R.

Bibliography

A, Abadie, and Imbens GW. "Large sample properties of matching estimators for average treatment effects." *Econometrica*, 2006: 74: 235–267.

An, Weihua. "Bayesian Propensity Score Estimators: Incorporating Uncertainties in Propensity Scores into Causal Inference." *Sociological Methodology*, 8 2010: 40(1): 151-189.

Austin, Peter C. "A comparison of 12 algorithms for matching on the propensity score." *Stat Med*, 3 15, 2014: 33(6):1057-69.

Austin, Peter C. "Optimal caliper widths for propensity-score matching when estimating differences in means and differences in proportions in observational studies." *Pharmaceut. Statist.*, 2011: 10: 150–161.

C, Bower, and Hansen M. "Assisted reproductive technologies and birth outcomes: overview of recent systematic reviews." *Reprod Fertil Dev.* , 2005: 17(3):329-33.

C, Drake . "Effects of misspecification of the propensity score on estimators of treatment effects." *Biometrics*, 1993: 49:1231–1236.

CA, Flores, and Mitnik OA. "Evaluating nonexperimental estimators for multiple." *Mimeo*, 2009.

Christopher , Winship, and Stephen Morgan. "The estimation of causal effects from observational data." *Annual Review of Sociology*, 1999: 25: 659–707.

CS, Armstrong, Jagolinzer AD, and Larcker D. "Chief Executive Officer Equity Incentives and Accounting Irregularities." *Journal of Accounting Research*, 2009: 48: 225–271.

DB, Rubin . "Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies." *Journal of Educational Psychology*, 1974: 66 (5), 688–701.

DB, Rubin. "Estimating causal effects from large data sets using propensity scores." *Ann Intern Med*, 10 15, 1997: 127(8 Pt 2):757-63.

—. "The use of propensity scores in applied Bayesian inference." *Bayesian Statistics*, 1985: 2: 463–472.

—. "The use of matched sampling and regression adjustment to remove bias in observational studies." *Biometrics* , 1973: 29: 185–203 .

—. "Bias Reduction Using Mahalanobis-Metric Matching." *Biometrics*, 1980: 36 (2): 293–298.

—. "Assignment to treatment group on the basis of a covariate." *Journal of Educational Statistics*, 1977: 2: 1-26.

—. "Using multivariate matched sampling and regression adjustment to control bias in observational studies." *Journal of the American Statistical Association* , 1979: 74: 318–328.

DB, Rubin, and Thomas N. "Matching using estimated propensity scores, relating theory to practice." *Biometrics*, 1996: 52: 249-264.

DB, Rubin, and Thomas N. "Characterizing the effect of matching using linear propensity score methods with normal distributions." *Biometrika*, 1992: 79: 797–809.

DE, Ho. "Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference." *Political Analysis*, 2007: 15: 199-236. .

DO, SCHARFSTEIN, ROTNITZKY A, and ROBINS JM. "Adjusting for non-ignorable drop-out using semiparametric non-response models." *J. Amer. Statist. Assoc.*, 1999: 94: 1096–1120.

Dong, Nianbo. "Using Propensity Score Methods to Approximate Factorial Experimental Designs to Analyze the Relationship between Two Variables and an Outcome." *American Journal of Evaluation*, 2015: 42-66.

DR, Cox, and Oakes D. *Analysis of Survival Data.* New York: Chapman & Hall., 1984.

E, Johnson, Dominici F, Griswold M, and Zeger S. "Disease Cases and Their Medical Costs Attributable to Smoking: An Analysis of the National Medical Expenditure Survey." *Journal of Econometrics*, 2003: 112: 135-151.

EA, Stuart . "Matching methods for causal inference: A review and a look forward." *Statistical science : a review journal of the Institute of Mathematical Statistics*, 2010: 25(1):1-21. doi:10.1214/09-STS313.

EA, STUART. "Developing practical recommendations for the use of propensity scores: Discussion of “A critical appraisal of propensity score matching in the medical

literature between 1996 and 2003” ” by P. Austin." *Stat. Med.*, 2008: 27: 2062–2065.

EA, Stuart. "Matching Methods for Causal Inference: A review and a look forward." *Statistical Science*, 2010: 25(1): 1-21.

G, Hong , and Raudenbush SW. "Evaluating kindergarten retention policy: A case study of causal inference for." *Journal of the American Statistical Association*, 2006: 101(475):901–910.

Guand, X, and Rosenbaum PR . "Comparison of multivariate matching methods: structures." *Journal of Computational and Graphical Statistics*, 1993: 2: 405–420.

GW , Imbens . "Nonparametric estimation of average treatment effects under exogeneity: a review." *Review of Economics and Statistics* , 2004: 86(1):4–29.

GW, Imbens. "The role of the propensity score in estimating dose-response functions." *Biometrika*, 2000: 87:706–710.

—. "Nonparametric estimation of average treatment effects under exogeneity: a review." *Review of Economics and Statistics* , 2004: 86(1): 4–29 .

GW, Snedecor, and Cochran WG. *Statistical Methods, Seventh Edition.* Iowa State University Press, 1980.

H, Smith. "Matching with multiple controls to estimate treatment effects in observational studies." *Sociological Methodology* , 1997: 27:325–353.

Harding, David J. "Counterfactual Models of Neighborhood Effects: The Effect of Neighborhood Poverty on Dropping Out and Teenage Pregnancy." *American Journal of Sociology*, 2003: 109: 676-719.

Huppler , Hullsiek K, and Louis TA. "Propensity score modeling strategies for the causal analysis of observational data." *Biostatistics*, 2002: 4:179-193.

J, COHEN. *Statistical Power Analysis for the Behavioral Science, 2nd ed.* Hillsdale, NJ.: Earlbaum, 1988.

J, HILL, Reiter J, and ZANUTTO E. *A comparison of experimental and observational data analyses. In Applied Bayesian Modeling and Causal Inference From an Incomplete Data Perspective.* Hoboken, NJ.: Wiley, 2004.

J, Hill, Rubin DB, and Thomas N. *The design of the New York School Choice Scholarship Program evaluation. In Research Designs: Inspired by the Work of Donald Campbell.* Thousand Oaks, CA: Sage, 1999.

J, ROBINS, and ROTNITZKY A. "Semiparametric efficiency in multivariate regression models with missing data." *J. Amer. Statist. Assoc.*, 1995: 90: 122–129.

J.M., Robins, Brumback B., and Hern´an M.A. "Marginal structural models and causal inference in epidemiology." *Epidemiology*, 2000: 11: 550–560.

JA, Cole, Loughlin JE, Ajene AN, Rosenberg DM, Cook SE, and Walker AM. "The effect of zanamivir treatment on influenza complications: a retrospective cohort study." *Clinical Therapeutics*, 2002: 24: 1824-1839.

JA, Hall, Summers KH, and Obenchain RL. "Cost and utilization comparisons among propensity score-matched insulin lispro and regular insulin users." *Journal of Managed Care Pharmacy*, 2003: 9: 263–268.

JD, Seeger, Walker AM, Williams PL, Saperia GM, and Sacks FM. "A propensity score matched cohort study of the effect of statins, mainly fluvastatin, on the occurrence of acute myocardial infarction." *American Journal of Cardiology*, 2003: 92: 1447–1451.

JJ, Heckman, Hidehiko H JJ, and Todd P. "Matching as an econometric evaluation estimator: Evidence from evaluating a job training programm." *Rev. Econom. Stud.* , 1997: 64: 605–654.

JL, Schafer, and Kang JD. "Average causal effects from nonrandomized studies: A practical guide and simulated case study." *Psychological Methods*, 2008: 13: 279–313.

JP, Weiss, Saynina O, McDonald KM, McClellan MB, and Hlatky MA. "Effectiveness and cost-effectiveness of implantable cardioverter defibrillators in the treatment of ventricular arrhythmias among Medicare beneficiaries." *The American Journal of Medicine*, 2002: 112: 519–527.

JZ, Ayanian, Landrum MB, and Guadagnoli E. "Specialty of ambulatory care physicians and mortality among elderly patients after myocardial infarction." *New England Journal of Medicine*, 2002: 347: 1678-1686.

K, Hirano, and Imbens GW. "The propensity score with continuous treatments." In *Applied Bayesian Modeling and Causal Inference from Incomplete-Data Perspectives: An Essential Journey with Donald Rubin's Statistical Family*, by Andrew Gelman, & Xiao-Li Meng, chap. 7. Wiley, 2004.

K, Hirano, Imbens GW, and Ridder G. "Efficient estimation of average treatment effects using the estimated propensity score." *Econometrica*, 2003: 71(4): 1307–1338.

Kosuke, Imai , and van Dyk DA. "Causal inference with general treatment regimes: Generalizing the propensity score." *Journal of the American Statistical Association* , 2004: 99(467):854–866.

Kosuke , Imai , and Ratkovic Marc . "Covariate balancing propensity score." *J. R. Statist. Soc. B*, 2014: 76(1): 243–263.

Kosuke, Imai. "Do Get-Out-The-Vote Calls Reduce Turnout? The Importance of Statistical Methods for Field Experiments." *American Political Science Review*, 2005: 99 (2): 283-300.

L, Mark. "Selecting an Appropriate Caliper Can Be Essential for Achieving Good Balance With Propensity Score Matching." *American Journal of Epidemiology*, 2013: 179: 226-235.

Lu B, Zanuto E, Hornik R, and Rosenbaum PR. "Matching with doses in an observational study of a media campaign against drug abuse." *Journal of the American Statistical Association*, 2001: 96:1245–1253.

M, Kendall. "A New Measure of Rank Correlation." *Biometrika*, 1938: 0 (1–2): 81–89.

M, Kewei, and Rosenbaum PR. "Substantial Gains in Bias Reduction from Matching with a Variable Number of Controls." *BIOMETRIC*, 3 2000: 56: 118-124.

MA, Brookhart, Schneeweiss S, Rothman KJ, Glynn RJ, Avorn J, and Sturmer T. "Variable selection for propensity score models." American Journal of Epidemiology, 2006: 163: 1149–1156.

McCandless, Lawrence C, Paul Gustafson, and Austin PC. "Bayesian Propensity Methods for Score Analysis for Observational Data." *Statistics in Medicine* , 2009: 28(1): 94-112.

MD, Larsen. "An Analysis of Survey Data on Smoking Using Propensity Scores." *Sankhya, Ser. B*, 1999: 61: 91–105.

ME, Sobel. "What do randomized studies of housing mobility demonstrate?: Causal inference in the face of interference." *Journal of the American Statistical Association*, 2006: 101(476):1398–1407.

MJ, Magee, Coombs LP, Peterson ED, and Mack MJ. "Patient selection and current practice strategy for off-pump coronary artery bypass surgery." *Circulation* , 2003: 108(S1): II9–14.

MM, Joffe, and Rosenbaum PR. "Propensity scores." *American Journal of Epidemiology*, 1999: 150:327–333.

NA, Christakis, and Iwashyna TJ. "The health impact of health care on families: a matched cohort study of hospice use by decedents and mortality outcomes in surviving, widowed spouses." *Social Science & Medicine*, 2003: 57: 465–475.

PC, Austin . "Optimal caliper widths for propensity-score matching when estimating differences in means and differences in proportions in observational studies." *Pharm. Stat.*, 2011: 10: 150–161.

PC, Austin. "Some methods of propensity-score matching had superior performance to others: results of an empirical investigation and Monte Carlo simulations." *Biometrical Journal*, 2009: 51(1): 171–184.

PC, Austin. "The performance of different propensity score methods for estimating marginal odds ratios." *Stat. Med.*, 2007: 26: 3078–3094.

PC, Austin, Grootendorst P, and Anderson GM. "A comparison of the ability of different propensity score models to balance measured variables between treated and untreated subjects: a Monte Carlo study." *Statist. Med.*, 2007: 26: 734–753.

Peter , McCullagh. "Regression Models for Ordinal Data." *Journal of the Royal Statistical Society. Series B (Methodological)*, 1980: 42: 109-142.

PR , Rosenbaum, and Rubin DB . "The central role of the propensity score in observational studies for causal effects." *Biometrika*, 1983: 70: 41-55.

PR, Rosenbaum. *Observational Studies. 2.* . New York, NY: Springer Verlag, 2002.

—. "The role of a second control group in an observational study." *Statistical Science* , 1987: 2(3):292–316.

RH , Dehejia , and Wahba S. "Causal effects in nonexperimental studies: Re-evaluating the evaluation of training programs." *Journal of the American Statistical Association*, 1999: 94:1053–62.

RH, Dehejia, and Wahba S. "Propensity score matching methods for non-experimental causal studies." *Review of Economics and Statistics*, 2002: 84:151–161.

Ronald, Fisher A. *The Design of Experiments (9th ed.).* Macmillan, 1971.

Ruth, Heller, Small Dylan , and Rosenbaum PR. "Split Samples and Design Sensitivity in Observational Studies." *Journal of the American Statistical Association*, 2009: 104 (487): 1090 - 1101.

S, Zeger, Wyant T, Miller L, and Samet J. "Statistical Testimony on Damages in Minnesota versus the Tobacco Industry." In *Statistical Science in the Courtroom, ed.*, by Gastwirth JL, 303–320. New York: Springer-Verlag, 2000.

Stephen , Morgan L, and David J Harding. "Matching Estimators of Causal Effects: Prospects and Pitfalls in Theory and Practice." *Sociological Methods & Research*, 2006: 35 (1): 3-60.

Tyler J. VanderWeele and Mirjam J. Knol "A Tutorial on Interaction" *Epidemiologic Methods* , 2014: 3 (1): 33–72.

TB, Ferguson, Coombs Jr LP, and Peterson E. "Internal thoracic artery grafting in the elderly patient undergoing coronary artery bypass grafting: room for process improvement? ." *Journal of Thoracic and Cardiovascular Surgery*, 2002: 123: 869–880.

Thomas, Diprete A, and Henriette Engelhardt. "Estimating Causal Effects With Matching Methods in the Presence and Absence of Bias Cancellation." *Sociological Metods & Research*, 2004: 32 (4): 501–528.

TJ, Iwashyna, and Lamont EB. "Effectiveness of adjuvant fluorouracil in clinical practice: a population population-based cohort study of elderly patients with stage III colon cancer." *Journal of Clinical Oncology*, 2002: 20: 3992-3998.

W, Raynor J. "Caliper Pair-Matching on a Continuous Variable in Case Control Studies." *Communications in Statistics: Theory and Methods*, 1983: 12:13, 1499-1509.

WG, Cochran , and Rubin DB. "Controlling bias in observational studies: A review." *Sankhya: The Indian Journal of Statistics, Series A*, 1973: 35:417–446.

WG, Cochran, and Chambers Paul S. "The Planning of Observational Studies of Human Populations." *Journal of the Royal Statistical Society. Series A (General)*, 1965: 234-266.

Zhao, Shandong, van Dyk A. David , and Imai Kosuke. "Causal Inference in Observational Studies with Non-Binary Treatments." *ARXIV*, 9 2013: 1309

APPENDIX: SAS Code

/*This Macro is used to conduct Monte Carlo Simulations for Chapter 3 and 4*/

```
libname mod "/home/zgu/TLO3/";
run;


filename junk dummy;
proc printto  log=junk; run;
%macro getmatch_nore(input,mnum,output);

data inputdata;set &input;run;

proc sql noprint;
select count(distinct(idt))
into :N_t
from inputdata;
quit;

%do %until(&N_t=0);


data tempodata;set inputdata;
if _N_=1;
run;


proc sql noprint;
select count(idt) into: countidt
from inputdata
where idt eq (select idt from tempodata);
quit;


%if  &countidt<=&mnum %then %do;
%let  ture_m=&countidt;
%end;


%if  &countidt>&mnum %then %do;
%let  ture_m=&mnum;
%end;

data rdata_tt;set inputdata;
if  _n_<=%eval(&ture_m);


proc sql noprint;
  create table inputdata2 as
  select *
```

```
  from inputdata a
  where a.idt not in (select idt FROM rdata_tt) and a.idc1 not in (select
idc1 FROM rdata_tt) and a.idc2 not in (select idc2 FROM rdata_tt) and
a.idc3 not in (select idc3 FROM rdata_tt);
quit;

data inputdata;set inputdata2;
run;


proc sql noprint;
select count(distinct(idt))
into :N_t
from inputdata;
quit;

data &output;set &output rdata_tt;
run;
%end;
%mend;



%macro post_process(input_data,index,out_data);
proc sql noprint;
create table C1_&index
as select distinct(idc1) as hid, 1/count(*) as weight
from &input_data
group by idt,idc1;
quit;


proc sql noprint;
create table C1_2&index
as select distinct(hid) as hid, sum(weight) as weight
from C1_&index
group by hid;
quit;

proc sql noprint;
create table C2_&index
as select distinct(idc2) as hid, 1/count(*) as weight
from &input_data
group by idt,idc2;
quit;


proc sql noprint;
create table C2_2&index
as select distinct(hid) as hid, sum(weight) as weight
from C2_&index
```

```
group by hid;
quit;



proc sql noprint;
create table C3_&index
as select distinct(idc3) as hid, 1/count(*) as weight
from &input_data
group by idt,idc3;
quit;



proc sql noprint;
create table C3_2&index
as select distinct(hid) as hid, sum(weight) as weight
from C3_&index
group by hid;
quit;




proc sql noprint;
create table treat&index
as select distinct(idt) as hid
from &input_data
group by idt;
quit;

data treat&index;set treat&index;
weight=1;
run;

data outdda;
set treat&index C1_2&index C2_2&index C3_2&index;



proc sql noprint;
  create table &out_data as
  select *
  from outdda a, temp b
  where a.hid=b.hid;
quit;

%mend;
```

```
%macro post_process_nore(input_data,index,out_data);


proc sql noprint;
create table treat&index
as select distinct(idt) as hid, count(*) as weight
from &input_data
group by idt;
quit;



proc sql noprint;
create table C1_&index
as select distinct(idc1) as hid, count(*) as weight
from &input_data
group by idc1;
quit;



proc sql noprint;
create table C2_&index
as select distinct(idc2) as hid, count(*) as weight
from &input_data
group by idc2;
quit;

proc sql noprint;
create table C3_&index
as select distinct(idc3) as hid, count(*) as weight
from &input_data
group by idc3;
quit;


data outdda;set treat&index C1_&index C2_&index C3_&index;
run;

proc sql;
  create table &out_data as
  select *
  from outdda a, temp b
  where a.hid=b.hid;
quit;

%mend;
```

```
%MACRO
MULTI_GROUP_MATCHING_B(N_TOTAL,Caliper1,PROPORTION_T,PROPORTION_M,REPLICA
TION,index);


data Gu_trimmed_OLS_Solu ;
set _null_;

data Gu_trimmed_ANOVA_Solu ;
set _null_;

/*
data  Solu_total_T_summary ;
set _null_;
data Solu_total_M_summary;
set _null_;

data  Solu_total_summary ;
set _null_;
data Solu_total_all;
set _null_;
RUN;
*/
data Solu_total_all;
set _null_;
RUN;
%let c_all=%eval(0);
data BIMA1;
SET _null_;
data BIMA2;
SET _null_;
data BIMA3;
SET _null_;
data BIMA4;
SET _null_;
data BIMA5;
SET _null_;


data Solu10_Gu_OLS;
set _null_;

data Solu10_Gu_ANOVA;
set _null_;

data So_STRA_7cdOLS;
set _null_;

data So_STRA_7cdANOVA;
set _null_;
```

```
data So_mu_matOLS;
set _null_;

data So_mu_matANOVA;
set _null_;

data Solu12_Gu_OLS;
set _null_;

data Solu12_Gu_ANOVA;
set _null_;
run;
```

```
%let Caliper_renew=%SYSEVALF(&Caliper1);

/*Replication Loop*/
%DO i=1 %to &REPLICATION;
dm "log;clear;odsresults; clear;";


%let Caliper1=&Caliper_renew;
```

```
data temp1;
do hid = 1 to  &N_TOTAL;
/*** Create Dataset ***/
/* Factor A -> T; Factor B -> M */

W1 = rannor(-1);
W2 = rannor(-1);
W3 = rannor(-1);
W4 = rannor(-1);
W5 = rannor(-1);
w6 = rannor(-1);
W7 = rannor(-1);
W8 = rannor(-1);
W9 = rannor(-1);
W10 = rannor(-1);
W11 = rannor(-1);
W12 = rannor(-1);
W13 = rannor(-1);
W14 = rannor(-1);
W15 = rannor(-1);
```

```
W16 = rannor(-1);



T1=1;
T=1;
M1=1;
M=1;
output;
end;
run;



proc iml;

use temp1;
   read all var _ALL_;
close temp1;
a=-5; b=5;


do until (sum(T)/&N_TOTAL-&PROPORTION_T=0) ; /* max iterations */
c = (a+b)/2;

T1 =c+
log(1.25)*W1+log(1.25)*W2+log(1.5)*W3+log(1.5)*W4+log(1.75)*w5+log(1.75)*
W6+log(2)*W7+log(2)*W8;
pi1=exp(T1)/(1+exp(T1));


T=rand('Bernoulli',pi1);
d=sum(T)/&N_TOTAL-&PROPORTION_T;
if d>0 then b = c;
if d<-0 then a = c;
end;

a=-5; b=5;

do until (sum(M)/&N_TOTAL-&PROPORTION_M=0) ; /* max iterations */
c = (a+b)/2;


M1 =c+
log(1.25)*W9+log(1.25)*W10+log(1.5)*W11+log(1.5)*W12+log(1.75)*w13+log(1.
75)*W14+log(2)*W15+log(2)*W16;


pi2=exp(M1)/(1+exp(M1));

M=rand('Bernoulli',pi2);
d=sum(M)/&N_TOTAL-&PROPORTION_M;
```

```
if d>0 then b = c;
if d<-0 then a = c;
end;
```

```
create temp var {HID W1 W2 W3 W4 W5 W6 W7 W8 W9 W10 W11 W12 W13 W14 W15
W16 T M}; /** create data set **/
append;       /** write data in vectors **/
close temp; /** close the data set **/
/*PS2 2 3 5 6 9 10
  PS1 2 5 6 8 10
*/

data temp;
set temp;
e = 3*rannor(-1);
Z
=log(1.25)*(W1+W2+W9+W10)+log(1.5)*(W3+W4+W11+W12)+log(1.75)*(W5+W6+W13+W
14)+log(2)*(W7+W8+W15+W16)+T+M+T*M+e;
drop e ;
run;

/*+log(2)*w10**3*/
```

```
data temp;
set temp;
if t=1 and M=1 then TM_cell = 1;
if t=1 and M=0 then TM_cell = 2;
if t=0 and M=1 then TM_cell = 3;
if t=0 and M=0 then TM_cell = 4;
run;

/* In ANOVA framework compared with OLS:
proc glm data = PS_weight_adj;
class TM_cell;
model z = w1 w2 w3  TM_cell  /solution;
weight      PS_weight_adj;
run; quit;

Cell 4 as reference group = 0; Cell 3 = Beta(M); Cell 2 = Beta (T);
Cell 1 = Cell 3 + Cell 2 + Beta (M*T)
*/
```

```
/*** #6: multiple matching based on ONE propensity scores estimated from
multinomial model
 all matching to Cell 4 ***/


/* Calculate Propensity Score: version 2: two binary logistic models */



/*Match witout Replacement using T1M1 group as the core matching group*/

/*Gu 03/16/2015*/
*%include "C:\Users\Zirui Gu\Desktop\DISSERTATON\Dissertation Matching
Macro 08022015.sas";
/*Calculate PS1 and PS2 and output to PS1_DATA and PS2_DATA*/




%Macro propen1(dataset,policy,outdata);
ODS SELECT NONE;
PROC LOGISTIC DATA=&dataset Descend;
/*Class  &class ;*/
Model &policy = w1 w2 w3 w4 w5 w6 w7 w8 w9 w10 w11 w12 w13 w14 w15
w16/*&class */;
OUTPUT OUT= &outdata prob=prob ;
run;
ODS SELECT ALL;
%mend propen1;




%propen1(temp,T,PS1_DATA);
%propen1(temp,M,PS2_DATA);
DATA PS1_DATA;set PS1_DATA;rename prob=PS1;run;
DATA PS2_DATA;set PS2_DATA;rename prob=PS2;run;

/*
proc sort data=PS1_DATA;by hid;run;proc sort data=PS2_DATA;by hid;run;
data PS_all;merge PS1_DATA PS2_DATA;by hid;run;
*/

PROC SQL NOPRINT;
 create table PS_all as
 SELECT A.PS1,
 B.*
 FROM PS1_DATA A, PS2_DATA B
 WHERE A.HID=B.HID;
QUIT;

data PS_all2;set PS_all;run;
```

```
data PS_all2;set PS_all2;
PS1=log(PS1/(1-PS1));
PS2=log(PS2/(1-PS2));
run;

PROC STANDARD DATA=PS_all2 MEAN=0 STD=1 OUT=PS_all3;
  VAR PS1 PS2 ;
RUN;

data PS_all2;set PS_all3;run;

proc freq data=PS_all2 order=freq noprint;
tables TM_CELL/outpct out=groupfreq;
run;

data groupfreq;set groupfreq;
grouporder=_N_;
run;

proc sql noprint;
create table PS_all2 as select a.*, b.grouporder
from PS_all2 a, groupfreq b
where a.TM_CELL=b.TM_CELL;
quit;


data Trt;set PS_all2;where grouporder=4;
data C1;set PS_all2;where grouporder=3;
data C2;set PS_all2;where grouporder=2;
data C3;set PS_all2;where grouporder=1;
data Trt(keep=hid PS1 PS2);set Trt;run;
data Trt;SET Trt;
rename PS1=PS1T;rename PS2=PS2T;idt=hid;
data C1(keep=hid PS1 PS2);set C1;run;
data C1;set C1;
idc1=hid;rename PS1=PS1C1; rename PS2=PS2C1;run;
data C2(keep=hid PS1 PS2);set C2;run;
data C2;set C2;
idc2=hid;rename PS1=PS1C2; rename PS2=PS2C2;run;
data C3(keep=hid PS1 PS2);set C3; run;
data C3;set C3;
idc3=hid;rename PS1=PS1C3; rename PS2=PS2C3;run;



/**** Estimate Moderator Effect using STRATIFIED data based on TWO
propensity scores ****/

/**** Estimate Moderator Effect using STRATIFIED data based on TWO
propensity scores ****/
```

```
/**** #7c&7d: common support data: STRATA, OLS ****/

/*** Create 3 strata based on each propensity score (prob) (with label 0,
1,2)***/


/**** Estimate Moderator Effect using STRATIFIED data based on TWO
propensity scores ****/

/**** #7c&7d: common support data: STRATA, OLS ****/

/*** Create 3 strata based on each propensity score (prob) (with label 0,
1,2)***/

%propen1(temp,t,tmp_propt);
%propen1(temp,m,tmp_propm);

/* #7: trimmed data for common support */

data tmp_propt;
set tmp_propt;
rename prob = prob_on_t;
data tmp_propm;
set tmp_propm;
rename prob = prob_on_m;
run;



proc sql noprint;
create table tmp_proptm_comm
as select A.prob_on_t,B.*
from tmp_propt A,tmp_propm B
where A.hid=B.hid
order by B.hid;
quit;


/* propensity score on t and m */
/*
ODS SELECT NONE;
proc means data= tmp_proptm_comm;
class TM_cell ;
var prob_on_t prob_on_m;
ods output summary=PScore_range_tm;
run;
ODS SELECT ALL;
*/

PROC SQL noprint;
```

```
create table PScore_range_tm as
select TM_cell as TM_cell,
min(prob_on_m) as prob_on_m_Min,
max(prob_on_m) as prob_on_m_Max,
min(prob_on_t) as prob_on_t_Min,
max(prob_on_t) as prob_on_t_Max
from tmp_proptm_comm
Group by TM_cell;
quit;


data tmp_proptm_comm2;set tmp_proptm_comm;run;

PROC SQL NOPRINT;
  SELECT min(prob_on_m_Max) INTO :PScore_upper_m
  FROM PScore_range_tm;
  SELECT max(prob_on_m_Min) INTO :PScore_lower_m
  FROM PScore_range_tm;
  SELECT min(prob_on_t_Max) INTO :PScore_upper_t
  FROM PScore_range_tm;
  SELECT max(prob_on_t_Min) INTO :PScore_lower_t
  FROM PScore_range_tm;
  quit;

data tmp_proptm_comm;
set tmp_proptm_comm;
if prob_on_m < =&PScore_upper_m and prob_on_m > =&PScore_lower_m
and  prob_on_t <= &PScore_upper_t and prob_on_t >= &PScore_lower_t;
run;



data PS_STRATA_7cd; set tmp_proptm_comm;
      Strata_t=prob_on_t;Strata_m=prob_on_m;
proc rank data= PS_STRATA_7cd grouPS=3 descending out=PS_STRATA_7cd; var
Strata_t Strata_m;run;
data  PS_STRATA_7cd ;set PS_STRATA_7cd; strata_tm = 10*strata_t +
strata_m; run;
*prob is converted to stratum id;
proc sort data=PS_STRATA_7cd out=PS_STRATA_7cd; by strata_tm; run;


ODS SELECT NONE;
proc freq data = PS_STRATA_7cd;
table strata_tm;
ods output onewayfreqs=PS_STRATA_7cd_freq;
run;
ODS SELECT ALL;

data PS_STRATA_7cd_freq;
set PS_STRATA_7cd_freq;
sampling_weight = percent/100;
```

```
keep strata_tm  sampling_weight;
run;

ODS SELECT NONE;
proc glm data = PS_STRATA_7cd;
by strata_tm;
model z = w1 w2 w3 w4 w5 w6 w7 w8 w9 w10 w11 w12 w13 w14 w15 w16  t M t*M
/solution;
ods  output ParameterEstimates = PS_STRATA_7cd_OLS;
run;
ODS SELECT ALL;

data PS_STRATA_7cd_OLS ;
merge PS_STRATA_7cd_OLS  PS_STRATA_7cd_freq;
by strata_tm;
run;

ODS SELECT NONE;
proc means data =  PS_STRATA_7cd_OLS ;
class Parameter ;
weight      sampling_weight;
var estimate;
output out =  PS_STRATA_7cd_OLS2 Mean = ;
run;
ODS SELECT ALL;


data PS_STRATA_7cd_OLS2;
set PS_STRATA_7cd_OLS2;
if _TYPE_ ne 0;
drop  _TYPE_  _FREQ_ ;
Replication = &i;
Note2 = "7d.OLS:subclassification (Imai&van Dyk)";
run;


/**** TRIMMED data: STRATA, ANOVA ****/

ODS SELECT NONE;
proc glm data = PS_STRATA_7cd;
by strata_tm;
model z = t M t*M   /solution;
ods  output ParameterEstimates = PS_STRATA_7cd_ANOVA;
run; quit;
ODS SELECT ALL;


data PS_STRATA_7cd_ANOVA ;
merge PS_STRATA_7cd_ANOVA  PS_STRATA_7cd_freq;
by strata_tm;
run;
```

```
ODS SELECT NONE;
proc means data =  PS_STRATA_7cd_ANOVA ;
class Parameter ;
weight      sampling_weight;
var estimate;
output out =  PS_STRATA_7cd_ANOVA2 Mean = ;
run;
ODS SELECT ALL;


data PS_STRATA_7cd_ANOVA2;
set PS_STRATA_7cd_ANOVA2;
if _TYPE_ ne 0;
drop  _TYPE_  _FREQ_ ;
Replication = &i;
Note2 = "7c.ANOVA:subclassification (Imai&van Dyk)";
run;


data So_STRA_7cdOLS ;
length   Note2 $70 MODEL $5;
set So_STRA_7cdOLS
PS_STRATA_7cd_OLS2 ;
Note2 = "7d.OLS:subclassification (Imai&van Dyk)";
model='OLS';
run;

data So_STRA_7cdANOVA ;
length   Note2 $70 MODEL $5;
set So_STRA_7cdANOVA
PS_STRATA_7cd_ANOVA2 ;
Note2 = "7c.ANOVA:subclassification (Imai&van Dyk)";
model='ANOVA';
run;







/*IPTW*/


ODS SELECT NONE;
PROC LOGISTIC DATA=temp ;
class TM_cell;
Model TM_cell = w1 w2 w3 w4 w5 w6 w7 w8 w9 w10 w11 w12 w13 w14 w15 w16 /
link=glogit ;
OUTPUT OUT= PS_weight0 prob=prob ;
run;
```

```
ODS SELECT ALL;

/* #4a & 4b: full sample for IPTW */
/* Next a propensity score weight, also referred to as the
inverse probability of treatment weight (IPTW), is calculated as
the inverse of the propensity score.*/

data PS_weight;
set PS_weight0;
if TM_cell =_level_;
PS_weight=1/prob;
run;

/* As of now the weights are based on the entire study group and
would give more weight to the smaller treatment grouPS.
A SQL procedure creates a weight that reflects the sample size for
each of the treatment grouPS. */

proc sql;
create table PS_weight_adj as
select *, (count(*)/&N_TOTAL)*PS_weight as PS_weight_adj
from PS_weight
group by TM_cell;
quit;



/* #4c & 4d: full sample for OLS & ANCOVA: PScore as probability of being
in the same cell */

/*** Obtain sample with probability of being in Cell 4 ***/

data  PS_weight_un;
set PS_weight;
keep hid PS_weight;
run;
proc sql noprint;
create table PS_weight4
as select A.PS_weight,B.*
from PS_weight_un A,PS_weight0 B
where A.hid=B.hid and B._LEVEL_=4
order by B.hid;
quit;

/* #5: trimmed data for common support from #4c & 4d */

ODS SELECT NONE;
proc means data= PS_weight4;
class TM_cell ;
var prob;
ods output summary=PScore_range;
run;
```

```
ODS SELECT ALL;

PROC SQL NOPRINT;
  SELECT min(prob_Max) INTO :PScore_upper
  FROM PScore_range;
  SELECT max(prob_min) INTO :PScore_lower
  FROM PScore_range;
  quit;


data PS_weight_common;
set PS_weight4;
if prob < (&PScore_upper+0.001) and prob > (&PScore_lower-0.001);
run;

proc sql noprint;
select count(*)
into :N_common
from PS_weight_common;
quit;

proc sql noprint;
create table PS_weight_adj_common as
select *, (count(*)/&N_common)*PS_weight as PS_weight_adj
from PS_weight_common
group by TM_cell;
quit;
```

```
/*** #5: * Estimate Moderator Effect using inverse propensity score
weighting: trimmed data ****/
/* using simple weight, not adjusting for sample size */
ODS SELECT NONE;
proc glm data = PS_weight_adj_common;
model z = w1 w2 w3 w4 w5 w6 w7 w8 w9 w10 w11 w12 w13 w14 w15 w16  t M t*M
/solution;
weight      PS_weight;
ods  output ParameterEstimates =PS_weight_adj_com_solu;
run; quit;
ODS SELECT ALL;
```

```
ODS SELECT NONE;
proc glm data = PS_weight_adj_common;
model z = t M t*M   /solution;
weight      PS_weight;
ods  output ParameterEstimates = PS_weight_adj_com_ANOVA;
run; quit;
ODS SELECT ALL;


data PS_weight_adj_com_solu;
set PS_weight_adj_com_solu;
Replication = &i;
run;

data PS_weight_adj_com_ANOVA;
set PS_weight_adj_com_ANOVA;
Replication = &i;
run;


data So_mu_matOLS ;
length Note1 $70  Note2 $70 model $5;
set So_mu_matOLS
PS_weight_adj_com_solu ;
Note1 = "5.So_mu_matIPW_trimmed";
Note2 = "5b.So_mu_matIPW_trimmed_OLS:simple weight";
model='OLS';
run;

data So_mu_matANOVA ;
length Note1 $70  Note2 $70 model $5;
set So_mu_matANOVA
PS_weight_adj_com_ANOVA ;
Note1 = "5.So_mu_matIPW_trimmed";
Note2 = "5a.So_mu_matIPW_trimmed_ANOVA:simple weight";
model='ANOVA';
run;

/**** #7c&7d: common support data: STRATA, OLS ****/

/*** Create 3 strata based on each propensity score (prob) (with label 0,
1,2)***/

/*NN Caliper Width*/

PROC SQL noprint;
create table T_C1 as
 SELECT *
 FROM Trt, C1
order by idt ;
create table T_C2 as
 SELECT *
```

```
 FROM Trt, C2
order by idt ;
create table T_C3 as
 SELECT *
 FROM Trt, C3
order by idt ;
 create table C1_C2 as
 SELECT *
 FROM C1, C2;
 create table C1_C3 as
 SELECT *
 FROM C1, C3;
 create table C2_C3 as
 SELECT *
 FROM C2, C3;
 QUIT;

data T_C1;set T_C1;
d_T_C1=sqrt((PS1T-PS1C1)**2+(PS2T-PS2C1)**2);
RUN;

data T_C2;set T_C2;
d_T_C2=sqrt((PS1T-PS1C2)**2+(PS2T-PS2C2)**2);
RUN;

data T_C3;set T_C3;
d_T_C3=sqrt((PS1T-PS1C3)**2+(PS2T-PS2C3)**2);
RUN;

data C1_C2;set C1_C2;
d_C1_C2=sqrt((PS1C1-PS1C2)**2+(PS2C1-PS2C2)**2);
RUN;

data C1_C3;set C1_C3;
d_C1_C3=sqrt((PS1C1-PS1C3)**2+(PS2C1-PS2C3)**2);
RUN;

data C2_C3;set C2_C3;
d_C2_C3=sqrt((PS1C2-PS1C3)**2+(PS2C2-PS2C3)**2);
RUN;


proc sql noprint;
SELECT std(d_T_C1) into: S_T_C1
from T_C1;
SELECT std(d_T_C2) into: S_T_C2
from T_C2;
SELECT std(d_T_C3) into: S_T_C3
from T_C3;
SELECT std(d_C1_C2) into: S_C1_C2
from C1_C2;
SELECT std(d_C1_C3) into: S_C1_C3
```

```
from C1_C3;
SELECT std(d_C2_C3) into: S_C2_C3
from C2_C3;
quit;


data  T_C1_1 T_C1_2 T_C1_3 T_C1_4 T_C1_5 T_C1_6 T_C1_7 T_C1_8 T_C1_9
T_C1_10
      T_C1_11 T_C1_12 T_C1_13 T_C1_14 T_C1_15 T_C1_16 T_C1_17 T_C1_18
T_C1_19 T_C1_20
        T_C1_21 T_C1_22 T_C1_23 T_C1_24 T_C1_25 T_C1_26 T_C1_27 T_C1_28
T_C1_29 T_C1_30
        T_C1_31 T_C1_32 T_C1_33 T_C1_34 T_C1_35 T_C1_36 T_C1_37 T_C1_38
T_C1_39 T_C1_40
        T_C1_41 T_C1_42 T_C1_43 T_C1_44 T_C1_45 T_C1_46 T_C1_47 T_C1_48
T_C1_49 T_C1_50
        T_C1_51 T_C1_52 T_C1_53 T_C1_54 T_C1_55 T_C1_56 T_C1_57 T_C1_58
T_C1_59;set T_C1;
if d_T_C1<=(0.15+0.05*0)*&S_T_C1 then output T_C1_1;
if d_T_C1<=(0.15+0.05*1)*&S_T_C1 then output T_C1_2;
if d_T_C1<=(0.15+0.05*2)*&S_T_C1 then output T_C1_3;
if d_T_C1<=(0.15+0.05*3)*&S_T_C1 then output T_C1_4;
if d_T_C1<=(0.15+0.05*4)*&S_T_C1 then output T_C1_5;
if d_T_C1<=(0.15+0.05*5)*&S_T_C1 then output T_C1_6;
if d_T_C1<=(0.15+0.05*6)*&S_T_C1 then output T_C1_7;
if d_T_C1<=(0.15+0.05*7)*&S_T_C1 then output T_C1_8;
if d_T_C1<=(0.15+0.05*8)*&S_T_C1 then output T_C1_9;
if d_T_C1<=(0.15+0.05*9)*&S_T_C1 then output T_C1_10;
if d_T_C1<=(0.15+0.05*10)*&S_T_C1 then output T_C1_11;
if d_T_C1<=(0.15+0.05*11)*&S_T_C1 then output T_C1_12;
if d_T_C1<=(0.15+0.05*12)*&S_T_C1 then output T_C1_13;
if d_T_C1<=(0.15+0.05*13)*&S_T_C1 then output T_C1_14;
if d_T_C1<=(0.15+0.05*14)*&S_T_C1 then output T_C1_15;
if d_T_C1<=(0.15+0.05*15)*&S_T_C1 then output T_C1_16;
if d_T_C1<=(0.15+0.05*16)*&S_T_C1 then output T_C1_17;
if d_T_C1<=(0.15+0.05*17)*&S_T_C1 then output T_C1_18;
if d_T_C1<=(0.15+0.05*18)*&S_T_C1 then output T_C1_19;
if d_T_C1<=(0.15+0.05*19)*&S_T_C1 then output T_C1_20;
if d_T_C1<=(0.15+0.05*20)*&S_T_C1 then output T_C1_21;
if d_T_C1<=(0.15+0.05*21)*&S_T_C1 then output T_C1_22;
if d_T_C1<=(0.15+0.05*22)*&S_T_C1 then output T_C1_23;
if d_T_C1<=(0.15+0.05*23)*&S_T_C1 then output T_C1_24;
if d_T_C1<=(0.15+0.05*24)*&S_T_C1 then output T_C1_25;
if d_T_C1<=(0.15+0.05*25)*&S_T_C1 then output T_C1_26;
if d_T_C1<=(0.15+0.05*26)*&S_T_C1 then output T_C1_27;
if d_T_C1<=(0.15+0.05*27)*&S_T_C1 then output T_C1_28;
if d_T_C1<=(0.15+0.05*28)*&S_T_C1 then output T_C1_29;
if d_T_C1<=(0.15+0.05*29)*&S_T_C1 then output T_C1_30;
if d_T_C1<=(0.15+0.05*30)*&S_T_C1 then output T_C1_31;
if d_T_C1<=(0.15+0.05*31)*&S_T_C1 then output T_C1_32;
if d_T_C1<=(0.15+0.05*32)*&S_T_C1 then output T_C1_33;
if d_T_C1<=(0.15+0.05*33)*&S_T_C1 then output T_C1_34;
```

```
if d_T_C1<=(0.15+0.05*34)*&S_T_C1 then output T_C1_35;
if d_T_C1<=(0.15+0.05*35)*&S_T_C1 then output T_C1_36;
if d_T_C1<=(0.15+0.05*36)*&S_T_C1 then output T_C1_37;
if d_T_C1<=(0.15+0.05*37)*&S_T_C1 then output T_C1_38;
if d_T_C1<=(0.15+0.05*38)*&S_T_C1 then output T_C1_39;
if d_T_C1<=(0.15+0.05*39)*&S_T_C1 then output T_C1_40;
if d_T_C1<=(0.15+0.05*40)*&S_T_C1 then output T_C1_41;
if d_T_C1<=(0.15+0.05*41)*&S_T_C1 then output T_C1_42;
if d_T_C1<=(0.15+0.05*42)*&S_T_C1 then output T_C1_43;
if d_T_C1<=(0.15+0.05*43)*&S_T_C1 then output T_C1_44;
if d_T_C1<=(0.15+0.05*44)*&S_T_C1 then output T_C1_45;
if d_T_C1<=(0.15+0.05*45)*&S_T_C1 then output T_C1_46;
if d_T_C1<=(0.15+0.05*46)*&S_T_C1 then output T_C1_47;
if d_T_C1<=(0.15+0.05*47)*&S_T_C1 then output T_C1_48;
if d_T_C1<=(0.15+0.05*48)*&S_T_C1 then output T_C1_49;
if d_T_C1<=(0.15+0.05*49)*&S_T_C1 then output T_C1_50;
if d_T_C1<=(0.15+0.05*50)*&S_T_C1 then output T_C1_51;
if d_T_C1<=(0.15+0.05*51)*&S_T_C1 then output T_C1_52;
if d_T_C1<=(0.15+0.05*52)*&S_T_C1 then output T_C1_53;
if d_T_C1<=(0.15+0.05*53)*&S_T_C1 then output T_C1_54;
if d_T_C1<=(0.15+0.05*54)*&S_T_C1 then output T_C1_55;
if d_T_C1<=(0.15+0.05*55)*&S_T_C1 then output T_C1_56;
if d_T_C1<=(0.15+0.05*56)*&S_T_C1 then output T_C1_57;
if d_T_C1<=(0.15+0.05*57)*&S_T_C1 then output T_C1_58;
if d_T_C1<=100*&S_T_C1 then output T_C1_59;
run;


data  T_C2_1 T_C2_2 T_C2_3 T_C2_4 T_C2_5 T_C2_6 T_C2_7 T_C2_8 T_C2_9
T_C2_10
      T_C2_11 T_C2_12 T_C2_13 T_C2_14 T_C2_15 T_C2_16 T_C2_17 T_C2_18
T_C2_19 T_C2_20
        T_C2_21 T_C2_22 T_C2_23 T_C2_24 T_C2_25 T_C2_26 T_C2_27 T_C2_28
T_C2_29 T_C2_30
        T_C2_31 T_C2_32 T_C2_33 T_C2_34 T_C2_35 T_C2_36 T_C2_37 T_C2_38
T_C2_39 T_C2_40
        T_C2_41 T_C2_42 T_C2_43 T_C2_44 T_C2_45 T_C2_46 T_C2_47 T_C2_48
T_C2_49 T_C2_50
        T_C2_51 T_C2_52 T_C2_53 T_C2_54 T_C2_55 T_C2_56 T_C2_57 T_C2_58
T_C2_59;set T_C2;
if d_T_C2<=(0.15+0.05*0)*&S_T_C2 then output T_C2_1;
if d_T_C2<=(0.15+0.05*1)*&S_T_C2 then output T_C2_2;
if d_T_C2<=(0.15+0.05*2)*&S_T_C2 then output T_C2_3;
if d_T_C2<=(0.15+0.05*3)*&S_T_C2 then output T_C2_4;
if d_T_C2<=(0.15+0.05*4)*&S_T_C2 then output T_C2_5;
if d_T_C2<=(0.15+0.05*5)*&S_T_C2 then output T_C2_6;
if d_T_C2<=(0.15+0.05*6)*&S_T_C2 then output T_C2_7;
if d_T_C2<=(0.15+0.05*7)*&S_T_C2 then output T_C2_8;
if d_T_C2<=(0.15+0.05*8)*&S_T_C2 then output T_C2_9;
if d_T_C2<=(0.15+0.05*9)*&S_T_C2 then output T_C2_10;
if d_T_C2<=(0.15+0.05*10)*&S_T_C2 then output T_C2_11;
if d_T_C2<=(0.15+0.05*11)*&S_T_C2 then output T_C2_12;
```

```
if d_T_C2<=(0.15+0.05*12)*&S_T_C2 then output T_C2_13;
if d_T_C2<=(0.15+0.05*13)*&S_T_C2 then output T_C2_14;
if d_T_C2<=(0.15+0.05*14)*&S_T_C2 then output T_C2_15;
if d_T_C2<=(0.15+0.05*15)*&S_T_C2 then output T_C2_16;
if d_T_C2<=(0.15+0.05*16)*&S_T_C2 then output T_C2_17;
if d_T_C2<=(0.15+0.05*17)*&S_T_C2 then output T_C2_18;
if d_T_C2<=(0.15+0.05*18)*&S_T_C2 then output T_C2_19;
if d_T_C2<=(0.15+0.05*19)*&S_T_C2 then output T_C2_20;
if d_T_C2<=(0.15+0.05*20)*&S_T_C2 then output T_C2_21;
if d_T_C2<=(0.15+0.05*21)*&S_T_C2 then output T_C2_22;
if d_T_C2<=(0.15+0.05*22)*&S_T_C2 then output T_C2_23;
if d_T_C2<=(0.15+0.05*23)*&S_T_C2 then output T_C2_24;
if d_T_C2<=(0.15+0.05*24)*&S_T_C2 then output T_C2_25;
if d_T_C2<=(0.15+0.05*25)*&S_T_C2 then output T_C2_26;
if d_T_C2<=(0.15+0.05*26)*&S_T_C2 then output T_C2_27;
if d_T_C2<=(0.15+0.05*27)*&S_T_C2 then output T_C2_28;
if d_T_C2<=(0.15+0.05*28)*&S_T_C2 then output T_C2_29;
if d_T_C2<=(0.15+0.05*29)*&S_T_C2 then output T_C2_30;
if d_T_C2<=(0.15+0.05*30)*&S_T_C2 then output T_C2_31;
if d_T_C2<=(0.15+0.05*31)*&S_T_C2 then output T_C2_32;
if d_T_C2<=(0.15+0.05*32)*&S_T_C2 then output T_C2_33;
if d_T_C2<=(0.15+0.05*33)*&S_T_C2 then output T_C2_34;
if d_T_C2<=(0.15+0.05*34)*&S_T_C2 then output T_C2_35;
if d_T_C2<=(0.15+0.05*35)*&S_T_C2 then output T_C2_36;
if d_T_C2<=(0.15+0.05*36)*&S_T_C2 then output T_C2_37;
if d_T_C2<=(0.15+0.05*37)*&S_T_C2 then output T_C2_38;
if d_T_C2<=(0.15+0.05*38)*&S_T_C2 then output T_C2_39;
if d_T_C2<=(0.15+0.05*39)*&S_T_C2 then output T_C2_40;
if d_T_C2<=(0.15+0.05*40)*&S_T_C2 then output T_C2_41;
if d_T_C2<=(0.15+0.05*41)*&S_T_C2 then output T_C2_42;
if d_T_C2<=(0.15+0.05*42)*&S_T_C2 then output T_C2_43;
if d_T_C2<=(0.15+0.05*43)*&S_T_C2 then output T_C2_44;
if d_T_C2<=(0.15+0.05*44)*&S_T_C2 then output T_C2_45;
if d_T_C2<=(0.15+0.05*45)*&S_T_C2 then output T_C2_46;
if d_T_C2<=(0.15+0.05*46)*&S_T_C2 then output T_C2_47;
if d_T_C2<=(0.15+0.05*47)*&S_T_C2 then output T_C2_48;
if d_T_C2<=(0.15+0.05*48)*&S_T_C2 then output T_C2_49;
if d_T_C2<=(0.15+0.05*49)*&S_T_C2 then output T_C2_50;
if d_T_C2<=(0.15+0.05*50)*&S_T_C2 then output T_C2_51;
if d_T_C2<=(0.15+0.05*51)*&S_T_C2 then output T_C2_52;
if d_T_C2<=(0.15+0.05*52)*&S_T_C2 then output T_C2_53;
if d_T_C2<=(0.15+0.05*53)*&S_T_C2 then output T_C2_54;
if d_T_C2<=(0.15+0.05*54)*&S_T_C2 then output T_C2_55;
if d_T_C2<=(0.15+0.05*55)*&S_T_C2 then output T_C2_56;
if d_T_C2<=(0.15+0.05*56)*&S_T_C2 then output T_C2_57;
if d_T_C2<=(0.15+0.05*57)*&S_T_C2 then output T_C2_58;
if d_T_C2<=100*&S_T_C2 then output T_C2_59;
run;


data  T_C3_1 T_C3_2 T_C3_3 T_C3_4 T_C3_5 T_C3_6 T_C3_7 T_C3_8 T_C3_9
T_C3_10
```

```
      T_C3_11 T_C3_12 T_C3_13 T_C3_14 T_C3_15 T_C3_16 T_C3_17 T_C3_18
T_C3_19 T_C3_20
        T_C3_21 T_C3_22 T_C3_23 T_C3_24 T_C3_25 T_C3_26 T_C3_27 T_C3_28
T_C3_29 T_C3_30
        T_C3_31 T_C3_32 T_C3_33 T_C3_34 T_C3_35 T_C3_36 T_C3_37 T_C3_38
T_C3_39 T_C3_40
        T_C3_41 T_C3_42 T_C3_43 T_C3_44 T_C3_45 T_C3_46 T_C3_47 T_C3_48
T_C3_49 T_C3_50
        T_C3_51 T_C3_52 T_C3_53 T_C3_54 T_C3_55 T_C3_56 T_C3_57 T_C3_58
T_C3_59;set T_C3;
if d_T_C3<=(0.15+0.05*0)*&S_T_C3 then output T_C3_1;
if d_T_C3<=(0.15+0.05*1)*&S_T_C3 then output T_C3_2;
if d_T_C3<=(0.15+0.05*2)*&S_T_C3 then output T_C3_3;
if d_T_C3<=(0.15+0.05*3)*&S_T_C3 then output T_C3_4;
if d_T_C3<=(0.15+0.05*4)*&S_T_C3 then output T_C3_5;
if d_T_C3<=(0.15+0.05*5)*&S_T_C3 then output T_C3_6;
if d_T_C3<=(0.15+0.05*6)*&S_T_C3 then output T_C3_7;
if d_T_C3<=(0.15+0.05*7)*&S_T_C3 then output T_C3_8;
if d_T_C3<=(0.15+0.05*8)*&S_T_C3 then output T_C3_9;
if d_T_C3<=(0.15+0.05*9)*&S_T_C3 then output T_C3_10;
if d_T_C3<=(0.15+0.05*10)*&S_T_C3 then output T_C3_11;
if d_T_C3<=(0.15+0.05*11)*&S_T_C3 then output T_C3_12;
if d_T_C3<=(0.15+0.05*12)*&S_T_C3 then output T_C3_13;
if d_T_C3<=(0.15+0.05*13)*&S_T_C3 then output T_C3_14;
if d_T_C3<=(0.15+0.05*14)*&S_T_C3 then output T_C3_15;
if d_T_C3<=(0.15+0.05*15)*&S_T_C3 then output T_C3_16;
if d_T_C3<=(0.15+0.05*16)*&S_T_C3 then output T_C3_17;
if d_T_C3<=(0.15+0.05*17)*&S_T_C3 then output T_C3_18;
if d_T_C3<=(0.15+0.05*18)*&S_T_C3 then output T_C3_19;
if d_T_C3<=(0.15+0.05*19)*&S_T_C3 then output T_C3_20;
if d_T_C3<=(0.15+0.05*20)*&S_T_C3 then output T_C3_21;
if d_T_C3<=(0.15+0.05*21)*&S_T_C3 then output T_C3_22;
if d_T_C3<=(0.15+0.05*22)*&S_T_C3 then output T_C3_23;
if d_T_C3<=(0.15+0.05*23)*&S_T_C3 then output T_C3_24;
if d_T_C3<=(0.15+0.05*24)*&S_T_C3 then output T_C3_25;
if d_T_C3<=(0.15+0.05*25)*&S_T_C3 then output T_C3_26;
if d_T_C3<=(0.15+0.05*26)*&S_T_C3 then output T_C3_27;
if d_T_C3<=(0.15+0.05*27)*&S_T_C3 then output T_C3_28;
if d_T_C3<=(0.15+0.05*28)*&S_T_C3 then output T_C3_29;
if d_T_C3<=(0.15+0.05*29)*&S_T_C3 then output T_C3_30;
if d_T_C3<=(0.15+0.05*30)*&S_T_C3 then output T_C3_31;
if d_T_C3<=(0.15+0.05*31)*&S_T_C3 then output T_C3_32;
if d_T_C3<=(0.15+0.05*32)*&S_T_C3 then output T_C3_33;
if d_T_C3<=(0.15+0.05*33)*&S_T_C3 then output T_C3_34;
if d_T_C3<=(0.15+0.05*34)*&S_T_C3 then output T_C3_35;
if d_T_C3<=(0.15+0.05*35)*&S_T_C3 then output T_C3_36;
if d_T_C3<=(0.15+0.05*36)*&S_T_C3 then output T_C3_37;
if d_T_C3<=(0.15+0.05*37)*&S_T_C3 then output T_C3_38;
if d_T_C3<=(0.15+0.05*38)*&S_T_C3 then output T_C3_39;
if d_T_C3<=(0.15+0.05*39)*&S_T_C3 then output T_C3_40;
if d_T_C3<=(0.15+0.05*40)*&S_T_C3 then output T_C3_41;
if d_T_C3<=(0.15+0.05*41)*&S_T_C3 then output T_C3_42;
```

```
if d_T_C3<=(0.15+0.05*42)*&S_T_C3 then output T_C3_43;
if d_T_C3<=(0.15+0.05*43)*&S_T_C3 then output T_C3_44;
if d_T_C3<=(0.15+0.05*44)*&S_T_C3 then output T_C3_45;
if d_T_C3<=(0.15+0.05*45)*&S_T_C3 then output T_C3_46;
if d_T_C3<=(0.15+0.05*46)*&S_T_C3 then output T_C3_47;
if d_T_C3<=(0.15+0.05*47)*&S_T_C3 then output T_C3_48;
if d_T_C3<=(0.15+0.05*48)*&S_T_C3 then output T_C3_49;
if d_T_C3<=(0.15+0.05*49)*&S_T_C3 then output T_C3_50;
if d_T_C3<=(0.15+0.05*50)*&S_T_C3 then output T_C3_51;
if d_T_C3<=(0.15+0.05*51)*&S_T_C3 then output T_C3_52;
if d_T_C3<=(0.15+0.05*52)*&S_T_C3 then output T_C3_53;
if d_T_C3<=(0.15+0.05*53)*&S_T_C3 then output T_C3_54;
if d_T_C3<=(0.15+0.05*54)*&S_T_C3 then output T_C3_55;
if d_T_C3<=(0.15+0.05*55)*&S_T_C3 then output T_C3_56;
if d_T_C3<=(0.15+0.05*56)*&S_T_C3 then output T_C3_57;
if d_T_C3<=(0.15+0.05*57)*&S_T_C3 then output T_C3_58;
if d_T_C3<=1000*&S_T_C3 then output T_C3_59;
run;



/*when k=59 nearest neigbor no caliper*/
%do k=1 %to 59;


data T_C1_C2_C3;merge T_C1_&k(in=t1) T_C2_&k(in=t2) T_C3_&k(in=t3);
by idt;
if t1 eq t2 eq t3;
run;



data T_C1_C2_C3;set T_C1_C2_C3;
d_C1C2=sqrt((PS1C1-PS1C2)**2+(PS2C1-PS2C2)**2);
d_C1C3=sqrt((PS1C1-PS1C3)**2+(PS2C1-PS2C3)**2);
d_C2C3=sqrt((PS1C2-PS1C3)**2+(PS2C2-PS2C3)**2);
run;


%if %eval(&k)<=58 %then %do;
proc sql noprint;
create table T_C1_C2_C3 as select *,
d_C1C2+d_C1C3+d_C2C3+d_T_C1+d_T_C2+d_T_C3 as Total_D
from T_C1_C2_C3
where d_C1C2<=(0.15+0.05*(&k-1))*&S_C1_C2 and d_C1C3<=(0.15+0.05*(&k-
1))*&S_C1_C3 and d_C2C3<=(0.15+0.05*(&k-1))*&S_C2_C3
order by idt,Total_D;
QUIT;
%end;
```

```
%if %eval(&k)=59 %then %do;
proc sql noprint;
create table T_C1_C2_C3 as select *,
d_C1C2+d_C1C3+d_C2C3+d_T_C1+d_T_C2+d_T_C3 as Total_D
from T_C1_C2_C3
order by idt,Total_D;
quit;
%end;
/*
data T_C1_C2_C3;set T_C1_C2_C3;
if d_C1C2<=(0.15+0.05*(&k-1))*&S_C1_C2 and d_C1C3<=(0.15+0.05*(&k-
1))*&S_C1_C3 and d_C2C3<=(0.15+0.05*(&k-1))*&S_C2_C3;
run;
*/


DATA T_C1_C2_C3;
 SET T_C1_C2_C3;
 BY idt;
 templast=lag(Total_D);
 if first.idt then templast=Total_D;
 RETAIN count;
 if first.idt then count=1;
 if Total_D ne templast then
 count=count+1;
run;

data T_C1C2C3_m1;set T_C1_C2_C3;
where count=1;
run;

data T_C1C2C3_m2;set T_C1_C2_C3;
where count<=2;
run;

data T_C1C2C3_m3;set T_C1_C2_C3;
where count<=3;
run;

data T_C1C2C3_m4;set T_C1_C2_C3;
where count<=4;
run;

data T_C1C2C3_m5;set T_C1_C2_C3;
where count<=5;
run;
```

```
proc sql noprint;
create table treatm1
as select distinct(idt) as hid, count(*) as weight
from T_C1C2C3_m1
group by idt;
quit;



proc sql noprint;
create table C1_m1
as select distinct(idc1) as hid, count(*) as weight
from T_C1C2C3_m1
group by idc1;
quit;



proc sql noprint;
create table C2_m1
as select distinct(idc2) as hid, count(*) as weight
from T_C1C2C3_m1
group by idc2;
quit;

proc sql noprint;
create table C3_m1
as select distinct(idc3) as hid, count(*) as weight
from T_C1C2C3_m1
group by idc3;
quit;


data BIMA1;set treatm1 C1_m1 C2_m1 C3_m1;
run;

proc sql;
  create table BIMA1 as
  select a.weight,b.*
  from BIMA1 a, temp b
  where a.hid=b.hid;
quit;


/*Done with m=1*/

%post_process(T_C1C2C3_m2,m2,BIMA2);
%post_process(T_C1C2C3_m3,m3,BIMA3);
%post_process(T_C1C2C3_m4,m4,BIMA4);
%post_process(T_C1C2C3_m5,m5,BIMA5);
/*Done with match with replacement*/
```

```
/*Match without replacement m=1 to 5*/


%do m=1 %to 5;

data BIMA_nore&m;set _null_;

%getmatch_nore(T_C1_C2_C3,&m,BIMA_nore&m);
%post_process_nore(BIMA_nore&m,m&m,BIMAnore&m);

%end;



/*Randomly select rannor(-1)*/



%do m=1 %to 5;

ODS SELECT NONE;
proc glm data = BIMA&m;
model z = w1 w2 w3 w4 w5 w6 w7 w8 w9 w10 w11 w12 w13 w14 w15 w16  t m t*m
/solution;
weight weight;
ods  output ParameterEstimates = Gu_trimmed_OLS_Solu;
run; quit;
ODS SELECT ALL;

ODS SELECT NONE;
proc glm data = BIMAnore&m;
model z = w1 w2 w3 w4 w5 w6 w7 w8 w9 w10 w11 w12 w13 w14 w15 w16  t m t*m
/solution;
weight weight;
ods  output ParameterEstimates = Gu_trimmed_OLS_Solunore1;
run; quit;
ODS SELECT ALL;




ODS SELECT NONE;
proc glm data = BIMA&m;
model z = t m t*m  /solution;
ods  output ParameterEstimates = Gu_trimmed_ANOVA_Solu;
weight weight;
run; quit;
ODS SELECT ALL;
```

```
ODS SELECT NONE;
proc glm data = BIMAnore&m;
model z = t m t*m  /solution;
ods  output ParameterEstimates = Gu_trimmed_ANOVA_Solunore1;
weight weight;
run; quit;
ODS SELECT ALL;
```

```
data Gu_trimmed_OLS_Solu;
set Gu_trimmed_OLS_Solu;
length model $5. replacement $30.  sort_by $30.;
Replication = &i;
Caliper= &Caliper1;
m=&m;
replacement='Replacement';
model='OLS';
sort_by='NA';
run;
```

```
data Gu_trimmed_OLS_Solunore1;
set Gu_trimmed_OLS_Solunore1;
length model $5. replacement $30.  sort_by $30.;
Replication = &i;
Caliper= &Caliper1;
m=&m;
replacement='Without Replacement';
model='OLS';
sort_by='Random';
run;
```

```
data Gu_trimmed_ANOVA_Solu;
set Gu_trimmed_ANOVA_Solu;
length model $5. replacement $30.  sort_by $30.;
Replication = &i;
Caliper= &Caliper1;
m=&m;
replacement='Replacement';
```

```
model='ANOVA';
sort_by='NA';
run;



data Gu_trimmed_ANOVA_Solunore1;
set Gu_trimmed_ANOVA_Solunore1;
length model $5. replacement $30.  sort_by $30.;
Replication = &i;
Caliper= &Caliper1;
m=&m;
replacement='Without Replacement';
model='ANOVA';
sort_by='Random';
run;



data Solu10_Gu_OLS;
length  Note1 $70;
set Solu10_Gu_OLS
Gu_trimmed_OLS_Solu Gu_trimmed_OLS_Solunore1  ;
Note1 = "NN Caliper Linear Regression";
run;

data Solu10_Gu_ANOVA;
length Note1 $70;
set Solu10_Gu_ANOVA
Gu_trimmed_ANOVA_Solu Gu_trimmed_ANOVA_Solunore1  ;
Note1 = "NN Caliper ANOVA";
run;

%end;
/*End of M from 1 to 5*/

%if %eval(&k)<=58 %then %do;
%let Caliper1=%SYSEVALF(&Caliper1+0.05);
%end;
%if %eval(&k)=59 %then %do;
%let Caliper1=%SYSEVALF(1000);
%end;

%end;
/*End of Caliper loop*/
/*End of Replications*/
```

```
proc freq data=tmp_proptm_comm order=freq noprint;
tables TM_CELL/outpct out=groupfreq;
run;

data groupfreq;set groupfreq;
grouporder=_N_;
run;

proc sql noprint;
create table tmp_proptm_comm as select a.*, b.grouporder
from tmp_proptm_comm a, groupfreq b
where a.TM_CELL=b.TM_CELL;
quit;


data Trt;set tmp_proptm_comm;where grouporder=4;
data C1;set tmp_proptm_comm;where grouporder=3;
data C2;set tmp_proptm_comm;where grouporder=2;
data C3;set tmp_proptm_comm;where grouporder=1;
data Trt(keep=hid PS1 PS2);set Trt;run;
data Trt;SET Trt;
rename PS1=PS1T;rename PS2=PS2T;idt=hid;
data C1(keep=hid PS1 PS2);set C1;run;
data C1;set C1;
idc1=hid;rename PS1=PS1C1; rename PS2=PS2C1;run;
data C2(keep=hid PS1 PS2);set C2;run;
data C2;set C2;
idc2=hid;rename PS1=PS1C2; rename PS2=PS2C2;run;
data C3(keep=hid PS1 PS2);set C3; run;
data C3;set C3;
idc3=hid;rename PS1=PS1C3; rename PS2=PS2C3;run;


data Trt;
set Trt;
RandomNumber= ranuni(12345);
output;
run;




/*NN Without Replacement*/


PROC SQL noprint;
create table T_C1 as
 SELECT *
```

```
 FROM Trt, C1
order by idt ;
create table T_C2 as
 SELECT *
 FROM Trt, C2
order by idt ;
create table T_C3 as
 SELECT *
 FROM Trt, C3
order by idt ;
 create table C1_C2 as
 SELECT *
 FROM C1, C2;
 create table C1_C3 as
 SELECT *
 FROM C1, C3;
 create table C2_C3 as
 SELECT *
 FROM C2, C3;
 QUIT;

data T_C1;set T_C1;
d_T_C1=sqrt((PS1T-PS1C1)**2+(PS2T-PS2C1)**2);
RUN;

data T_C2;set T_C2;
d_T_C2=sqrt((PS1T-PS1C2)**2+(PS2T-PS2C2)**2);
RUN;

data T_C3;set T_C3;
d_T_C3=sqrt((PS1T-PS1C3)**2+(PS2T-PS2C3)**2);
RUN;

data C1_C2;set C1_C2;
d_C1_C2=sqrt((PS1C1-PS1C2)**2+(PS2C1-PS2C2)**2);
RUN;

data C1_C3;set C1_C3;
d_C1_C3=sqrt((PS1C1-PS1C3)**2+(PS2C1-PS2C3)**2);
RUN;

data C2_C3;set C2_C3;
d_C2_C3=sqrt((PS1C2-PS1C3)**2+(PS2C2-PS2C3)**2);
RUN;


proc sql noprint;
SELECT std(d_T_C1) into: S_T_C1
from T_C1;
SELECT std(d_T_C2) into: S_T_C2
from T_C2;
SELECT std(d_T_C3) into: S_T_C3
```

```
from T_C3;
SELECT std(d_C1_C2) into: S_C1_C2
from C1_C2;
SELECT std(d_C1_C3) into: S_C1_C3
from C1_C3;
SELECT std(d_C2_C3) into: S_C2_C3
from C2_C3;
quit;

/*0.65 0.8 0.85 1.1 2.55*/



data  T_C1;set T_C1;
where d_T_C1<=(100)*&S_T_C1 ;
run;


data  T_C2;set T_C2;
where d_T_C2<=(100)*&S_T_C2 ;
run;


data  T_C3;set T_C3;
where d_T_C3<=(100)*&S_T_C3 ;
run;




data T_C1_C2_C3;merge T_C1(in=t1) T_C2(in=t2) T_C3(in=t3);
by idt;
if t1 eq t2 eq t3;
run;




data T_C1_C2_C3;set T_C1_C2_C3;
d_C1C2=sqrt((PS1C1-PS1C2)**2+(PS2C1-PS2C2)**2);
d_C1C3=sqrt((PS1C1-PS1C3)**2+(PS2C1-PS2C3)**2);
d_C2C3=sqrt((PS1C2-PS1C3)**2+(PS2C2-PS2C3)**2);
run;




proc sql noprint;
create table T_C1_C2_C3_2 as select *,
d_C1C2+d_C1C3+d_C2C3+d_T_C1+d_T_C2+d_T_C3 as Total_D
from T_C1_C2_C3
```

```
where d_C1C2<=100*&S_C1_C2 and d_C1C3<=100*&S_C1_C3 and
d_C2C3<=100*&S_C2_C3;
quit;


proc sql;
create table T_C1_C2_C3_2_2 as select a.*,b.RandomNumber from
T_C1_C2_C3_2 a, Trt b
where a.idt=b.hid
order by b.RandomNumber,a.Total_D;
quit;


/*
data T_C1_C2_C3;set T_C1_C2_C3;
if d_C1C2<=(0.15+0.05*(&k-1))*&S_C1_C2 and d_C1C3<=(0.15+0.05*(&k-
1))*&S_C1_C3 and d_C2C3<=(0.15+0.05*(&k-1))*&S_C2_C3;
run;

proc sql noprint;
create table T_C1_C2_C3_3 as select * from T_C1_C2_C3_2_2

quit;
*/



/*
proc sort data= T_C1_C2_C3;
by idt Total_D;
run;
*/



DATA T_C1C2C3_m1; set T_C1_C2_C3_2_2;
by  RandomNumber;
if first.RandomNumber then output;
run;



proc sql noprint;
create table treatm1
as select distinct(idt) as hid, count(idt) as weight
from T_C1C2C3_m1
group by idt;
quit;
```

```
proc sql noprint;
create table C1_m1
as select distinct(idc1) as hid, count(idc1) as weight
from T_C1C2C3_m1
group by idc1;
quit;



proc sql noprint;
create table C2_m1
as select distinct(idc2) as hid, count(idc2) as weight
from T_C1C2C3_m1
group by idc2;
quit;

proc sql noprint;
create table C3_m1
as select distinct(idc3) as hid, count(idc3) as weight
from T_C1C2C3_m1
group by idc3;
quit;


data BIMA1;set treatm1 C1_m1 C2_m1 C3_m1;
run;

proc sql;
  create table BIMA1 as
  select a.weight,b.*
  from BIMA1 a, temp b
  where a.hid=b.hid;
quit;




/*Match without replacement m=1 to 5*/


/*
data test;set T_C1_C2_C3_3; obs=_N_;
data test;set test;where obs<=10000;
run;
*/
/*T_C1_C2_C3_3*/


%getmatch_nore(T_C1_C2_C3,1,BIMA_nore1);
%post_process_nore(BIMA_nore1,m1,BIMA_nore2);
```

```
/*Randomly select rannor(-1)*/

ODS SELECT NONE;
proc glm data = BIMA_nore2;
model z = w1 w2 w3 w4 w5 w6 w7 w8 w9 w10 w11 w12 w13 w14 w15 w16  t m t*m
/solution;
ods  output ParameterEstimates = Gu_NN_OLS_Solu;
run;
ODS SELECT ALL;


ODS SELECT NONE;
proc glm data = BIMA_nore2;
model z = t m t*m  /solution;
ods  output ParameterEstimates = Gu_NN_ANOVA_Solu;
run;
ODS SELECT ALL;



ODS SELECT NONE;
proc glm data = BIMA1;
model z = w1 w2 w3 w4 w5 w6 w7 w8 w9 w10 w11 w12 w13 w14 w15 w16  t m t*m
/solution;
ods  output ParameterEstimates = Gu_NNre_OLS_Solu;
weight weight;
run;
ODS SELECT ALL;


ODS SELECT NONE;
proc glm data = BIMA1;
model z = t m t*m  /solution;
ods  output ParameterEstimates = Gu_NNre_ANOVA_Solu;
weight weight;
run;
ODS SELECT ALL;

data Gu_NN_OLS_Solu;
set Gu_trimmed_OLS_Solu;
length note2 $200. model $5;
model='OLS';
replication=&i;
note2='NN Without Replacement Linear Regression';
run;


data Gu_NNre_OLS_Solu;
set Gu_NNre_OLS_Solu;
```

```
length note2 $200. model $5;
model='OLS';
replication=&i;
note2='NN With Replacement Linear Regression';
run;


data Gu_NN_ANOVA_Solu;
set Gu_NN_ANOVA_Solu;
length note2 $200. model $5;
replication=&i;
model='ANOVA';
note2='NN Without Replacement ANOVA';
run;

data Gu_NNre_ANOVA_Solu;
set Gu_NNre_ANOVA_Solu;
length note2 $200. model $5;
replication=&i;
model='ANOVA';
note2='NN With Replacement ANOVA';
run;

data Solu12_Gu_OLS;set Solu12_Gu_OLS Gu_NN_OLS_Solu Gu_NNre_OLS_Solu;
m=1;
run;

data Solu12_Gu_ANOVA;set Solu12_Gu_ANOVA Gu_NN_ANOVA_Solu
Gu_NNre_ANOVA_Solu;
m=1;
run;




%end;



/*Estimate treatment effect using trimmed Matched data*/

data Solu_total_all2;
length note2 $200.;
set
Solu10_Gu_OLS Solu10_Gu_ANOVA;
note2=strip(note1)||"Caliper="||strip(caliper)||"m="||strip(m)||strip(sor
t_by)||strip(replacement);
run;

data Solu_total_all;set Solu_total_all2 So_STRA_7cdOLS
```

```
So_STRA_7cdANOVA
So_mu_matOLS
So_mu_matANOVA
Solu12_Gu_OLS
Solu12_Gu_ANOVA;
run;


data So_all_&index ;
set _null_;run;

data So_all_&index;
set So_all_&index Solu_total_all;
Squared_deviance = (Estimate - 1)*(Estimate - 1);
Bias = abs(Estimate - 1);
PROPORTION_T = &PROPORTION_T;
PROPORTION_M = &PROPORTION_M;
if parameter in ('T','M','T*M') then output;
run;



data mod.So_all_&index;set So_all_&index;
run;


%mend;

options nonotes;

%MULTI_GROUP_MATCHING_B(1000,0.15,0.2,0.2,10,1);
%MULTI_GROUP_MATCHING_B(1000,0.15,0.3,0.3,10,2);
%MULTI_GROUP_MATCHING_B(1000,0.15,0.4,0.4,10,3);
%MULTI_GROUP_MATCHING_B(1000,0.15,0.5,0.5,10,4);
%MULTI_GROUP_MATCHING_B(1000,0.15,0.6,0.6,10,5);
%MULTI_GROUP_MATCHING_B(1000,0.15,0.7,0.7,10,6);
%MULTI_GROUP_MATCHING_B(1000,0.15,0.8,0.8,10,7);
%MULTI_GROUP_MATCHING_B(1000,0.15,0.3,0.7,10,25);
%MULTI_GROUP_MATCHING_B(1000,0.15,0.3,0.2,10,21);
%MULTI_GROUP_MATCHING_B(1000,0.15,0.3,0.4,10,22);
%MULTI_GROUP_MATCHING_B(1000,0.15,0.3,0.5,10,23);
%MULTI_GROUP_MATCHING_B(1000,0.15,0.3,0.6,10,24);
%MULTI_GROUP_MATCHING_B(1000,0.15,0.3,0.8,10,26);
%MULTI_GROUP_MATCHING_B(1000,0.15,0.2,0.3,10,31);
%MULTI_GROUP_MATCHING_B(1000,0.15,0.2,0.4,10,32);
%MULTI_GROUP_MATCHING_B(1000,0.15,0.2,0.5,10,33);
%MULTI_GROUP_MATCHING_B(1000,0.15,0.2,0.6,10,34);
%MULTI_GROUP_MATCHING_B(1000,0.15,0.2,0.7,10,35);
%MULTI_GROUP_MATCHING_B(1000,0.15,0.2,0.8,10,36);
%MULTI_GROUP_MATCHING_B(1000,0.15,0.4,0.2,10,41);
%MULTI_GROUP_MATCHING_B(1000,0.15,0.4,0.3,10,42);
%MULTI_GROUP_MATCHING_B(1000,0.15,0.4,0.5,10,43);
```

```
%MULTI_GROUP_MATCHING_B(1000,0.15,0.4,0.6,10,44);
%MULTI_GROUP_MATCHING_B(1000,0.15,0.4,0.7,10,45);
%MULTI_GROUP_MATCHING_B(1000,0.15,0.4,0.8,10,46);
%MULTI_GROUP_MATCHING_B(1000,0.15,0.5,0.2,10,51);
%MULTI_GROUP_MATCHING_B(1000,0.15,0.5,0.3,10,52);
%MULTI_GROUP_MATCHING_B(1000,0.15,0.5,0.4,10,53);
%MULTI_GROUP_MATCHING_B(1000,0.15,0.5,0.6,10,54);
%MULTI_GROUP_MATCHING_B(1000,0.15,0.5,0.7,10,55);
%MULTI_GROUP_MATCHING_B(1000,0.15,0.5,0.8,10,56);

proc printto; run;
```

/*This part Produces Figures for Chapter 3*/

```
libname simu  'C:\Users\Zirui Gu\Desktop\test\result\';

%macro plotchap3_eq(datain,name1,name2,name3,name4,name5,name6);

Legend1 label=(height=1 position=top justify=center
 '')
 value=(h=1.5 'GIPTW' 'NNWR' 'NN'  '3X3 SP')
 across=1 down=4
 position = (top left inside)
 mode=protect
 offset = (3 pct);


axis11 label=(a=90 h=7 pct ' ') value=(h=1) order=(0 to 3 by 0.5)
offset=(1,1);
axis12 label=(a=90 h=7 pct ' ') value=(h=1) order=(0 to 3 by 0.5)
offset=(1,1);
axis13 label=(a=90 h=7 pct ' ') value=(h=1) order=(0 to 3 by 0.5)
offset=(1,1);
axis14 label=(h=1.2 pct ' ') offset=(1,1) value=(h=1);

axis21 label=(a=90 h=7 pct ' ') value=(h=1) order=(0 to 0.3 by 0.05)
offset=(1,1);
axis22 label=(a=90 h=7 pct ' ') value=(h=1) order=(0 to 0.3 by 0.05)
offset=(1,1);
axis23 label=(a=90 h=7 pct ' ') value=(h=1) order=(0 to 0.3 by 0.05)
offset=(1,1);
axis24 label=(h=1.2 pct ' ')offset=(1,1) value=(h=1);

/*

axis11 label=(h=1.2 angle=90  "MSE T1xT2") value=(h=1) order=(0 to 3 by
0.5) offset=(1,1);
```

```
axis12 label=(h=1.2 angle=90  "MSE T1") value=(h=1) order=(0 to 3 by 0.5)
offset=(1,1);
axis13 label=(h=1.2 angle=90  "MSE T2") value=(h=1) order=(0 to 3 by 0.5)
offset=(1,1);
axis14 label=(h=1.2 "Prevalence of T1,T2")value=(h=1);

axis21 label=(h=1.2 angle=90  "RB T1*T2") value=(h=1) order=(0 to 0.3 by
0.05) offset=(1,1);
axis22 label=(h=1.2 angle=90  "RB T1") value=(h=1) order=(0 to 0.3 by
0.05) offset=(1,1);
axis23 label=(h=1.2 angle=90  "RB T2") value=(h=1) order=(0 to 0.3 by
0.05) offset=(1,1);
axis24 label=(h=1.2 "Prevalence of T1,T2")value=(h=1);
*/


symbol1 i=join w=1.5 line=1 v=dot color=black h=1;
symbol2 i=join w=1.5 line=2 v=dot  color=green h=1;
symbol3 i=join w=1.5 line=4 v=dot color=blue h=1;
symbol4 i=join w=1.5 line=20 v=dot color=red h=1;


/*
MSE_Dif_Impact_M1_M0
MSE_Impact_on_M0
MSE_Impact_on_T0

Bias_Dif_Impact_M1_M0
Bias_Impact_M0
Bias_Impact_T0
*/
proc gplot data=&datain;
plot MSE_Dif_Impact_M1_M0*proportion_TM=methods/name=&name1 vaxis= axis11
haxis=axis14 LEGEND=lengend1;
where Proportion_T eq Proportion_M;
note
f='Times New Roman' m=(8,40)pct h=1.2 a=90 'MSE('
f=greek  h=2 't'
f=simplex h=3  m=(+1.1,-1.6) "^"
f=ITALIC m=(-0.8,-0.9) h=1 'T'
f=greek m=(+.0, -.2) h=.5 '1'
f=simplex m=(-.0, +.2)  h=1 'x'
f=ITALIC  h=1 'T'
f=greek m=(+.0,-.2) h=.5 '2'
f='Times New Roman' m=(-0.3, +0.5) h=1.2   ')';
note
f='Times New Roman' m=(46,2.5)pct h=1.2 'Prevalance of'
f=ITALIC m=(+0,-0) h=1.2  'T'
f=greek  m=(-0.3,-0.2) h=0.8  '1'
f=ITALIC m=(+0.3,+0.2) h=1.2  ',T'
f=greek  m=(-0.3,-0.2) h=0.8  "2";
```

```
FOOTNOTE  m=(44,0)pct f='Times New Roman' h=1.2  '(a):MSE of ACIE
Estimates';
run;


proc gplot data=&datain;
plot Bias_Dif_Impact_M1_M0*proportion_TM=methods/name=&name2 vaxis=
axis21 haxis=axis24 LEGEND=lengend1;
where Proportion_T eq Proportion_M;
note
f='Times New Roman' m=(8,40)pct h=1.2 a=90 'RB('
f=greek  h=2 't'
f=simplex h=3  m=(+1.1,-1.6) "^"
f=ITALIC m=(-0.8,-0.9) h=1 'T'
f=greek m=(+.0, -.2) h=.5 '1'
f=simplex m=(-.0, +.2)  h=1 'x'
f=ITALIC  h=1 'T'
f=greek m=(+.0,-.2) h=.5 '2'
f='Times New Roman' m=(-0.3, +0.5) h=1.2   ')';
note
f='Times New Roman' m=(46,2.5)pct h=1.2 'Prevalance of'
f=ITALIC m=(+0,-0) h=1.2  'T'
f=greek  m=(-0.3,-0.2) h=0.8  '1'
f=ITALIC m=(+0.3,+0.2) h=1.2  ',T'
f=greek  m=(-0.3,-0.2) h=0.8  "2";
FOOTNOTE  m=(44,0)pct f='Times New Roman' h=1.2  '(b):RB of ACIE
Estimates';
run;



proc gplot data=&datain;
plot MSE_Impact_on_M0*proportion_TM=methods/name=&name3 vaxis= axis12
haxis=axis14 LEGEND=lengend1;
where Proportion_T eq Proportion_M;
note
f='Times New Roman' m=(8,40)pct h=1.2 a=90 'MSE('
f=greek  h=2 't'
f=simplex h=3  m=(+1.1,-1.6) "^"
f=ITALIC m=(-0.8,-0.9) h=1 'T'
f=greek m=(+0, -.2) h=.5 '1'
f='Times New Roman' m=(-0.3, +0.5) h=1.2   ')';
note
f='Times New Roman' m=(46,2.5)pct h=1.2 'Prevalance of'
f=ITALIC m=(+0,-0) h=1.2  'T'
f=greek  m=(-0.3,-0) h=0.8  '1'
f=ITALIC m=(+0.3,+0) h=1.2  ',T'
f=greek  m=(-0.3,-0) h=0.8  "2";
FOOTNOTE  m=(44,0)pct f='Times New Roman' h=1.2  '(c):MSE of ACME
Estimates for ' f=ITALIC m=(+0,-0) h=1.2  'T' m=(-0,-0) f=greek h=0.8
'1';
run;
```

```
proc gplot data=&datain;
plot Bias_Impact_M0*proportion_TM=methods/name=&name4 vaxis= axis22
haxis=axis24 LEGEND=lengend1;
where Proportion_T eq Proportion_M;
note
f='Times New Roman' m=(8,40)pct h=1.2 a=90 'RB('
f=greek  h=2 't'
f=simplex h=3  m=(+1.1,-1.6) "^"
f=ITALIC m=(-0.8,-0.9) h=1 'T'
f=greek m=(+0, -.2) h=.5 '1'
f='Times New Roman' m=(-0.3, +0.5) h=1.2   ')';
note
f='Times New Roman' m=(46,2.5)pct h=1.2 'Prevalance of'
f=ITALIC m=(+0,-0) h=1.2  'T'
f=greek  m=(-0.3,-0) h=0.8  '1'
f=ITALIC m=(+0.3,+0) h=1.2  ',T'
f=greek  m=(-0.3,-0) h=0.8  "2";
FOOTNOTE  m=(44,0)pct f='Times New Roman' h=1.2  '(d):RB of ACME
Estimates for ' f=ITALIC m=(+0,-0) h=1.2  'T' m=(-0,-0) f=greek h=0.8
'1';
run;

proc gplot data=&datain;
plot MSE_Impact_on_T0*proportion_TM=methods/name=&name5 vaxis= axis13
haxis=axis14 LEGEND=lengend1;
where Proportion_T eq Proportion_M;
note
f='Times New Roman' m=(8,40)pct h=1.2 a=90 'MSE('
f=greek  h=2 't'
f=simplex h=3  m=(+1.1,-1.6) "^"
f=ITALIC m=(-0.8,-0.9) h=1 'T'
f=greek m=(+0, -.2) h=.5 '2'
f='Times New Roman' m=(-0.3, +0.5) h=1.2   ')';
note
f='Times New Roman' m=(46,2.5)pct h=1.2 'Prevalance of'
f=ITALIC m=(+0,-0) h=1.2  'T'
f=greek  m=(-0.3,-0) h=0.8  '1'
f=ITALIC m=(+0.3,+0) h=1.2  ',T'
f=greek  m=(-0.3,-0) h=0.8  "2";
FOOTNOTE  m=(44,0)pct f='Times New Roman' h=1.2  '(e):MSE of ACME
Estimates for ' f=ITALIC m=(+0,-0) h=1.2  'T' m=(-0,-0) f=greek h=0.8
'2';
run;


proc gplot data=&datain;
plot Bias_Impact_T0*proportion_TM=methods/name=&name6 vaxis= axis23
haxis=axis24 LEGEND=lengend1;
where Proportion_T eq Proportion_M;
note
```

```
f='Times New Roman' m=(8,40)pct h=1.2 a=90 'RB('
f=greek  h=2 't'
f=simplex h=3  m=(+1.1,-1.6) "^"
f=ITALIC m=(-0.8,-0.9) h=1 'T'
f=greek m=(+0, -.2) h=.5 '2'
f='Times New Roman' m=(-0.3, +0.5) h=1.2   ')';
note
f='Times New Roman' m=(46,2.5)pct h=1.2 'Prevalance of'
f=ITALIC m=(+0,-0) h=1.2  'T'
f=greek  m=(-0.3,-0) h=0.8  '1'
f=ITALIC m=(+0.3,+0) h=1.2  ',T'
f=greek  m=(-0.3,-0) h=0.8  "2";
FOOTNOTE  m=(44,0)pct f='Times New Roman' h=1.2  '(f):RB of ACME
Estimates for ' f=ITALIC m=(+0,-0) h=1.2  'T' m=(-0,-0) f=greek h=0.8
'2';
run;

%mend;



%macro
plotchap3_ne(datain,p1,v1max,v2max,v3max,v4max,v5max,v6max,name1,name2,na
me3,name4,name5,name6);
/*
if note2 eq 'IPTW Linear Regression' then method='1.GIPTW within CS';
if note2 eq 'NN With Replacement Linear Regression' then method='2.NN
within CS';
if note2 eq 'NN Without Replacement Linear Regression' then
method='3.NNWR within CS';
if note2 eq 'NN Caliper Linear RegressionCaliper=3.05m=1RandomWithout
Replacement' then method='4.NN without CS';
if note2 eq 'NN Caliper Linear RegressionCaliper=3.05m=1NAReplacement'
then method='5.NNWR without CS';
if note2 eq 'Stratification on P-Function' then method='6.3X3 SP within
CS';
*/
Legend1 label=(height=1 position=top justify=center
 '')
 value=(h=1.2 'GIPTW' 'NNWR' 'NN'  '3X3 SP')
 across=1 down=4
 position = (top left inside)
 mode=protect
 offset = (3 pct);




axis11 label=(a=90 h=7 pct ' ') value=(h=1) order=(0 to &v1max by 0.5)
offset=(1,1);
```

```
axis12 label=(a=90 h=7 pct ' ') value=(h=1) order=(0 to &v2max by 0.5)
offset=(1,1);
axis13 label=(a=90 h=7 pct ' ') value=(h=1) order=(0 to &v3max by 0.5)
offset=(1,1);
axis14 label=(h=1.2 pct ' ') value=(h=1);

axis21 label=(a=90 h=7 pct ' ') value=(h=1) order=(0 to &v4max by 0.05)
offset=(1,1);
axis22 label=(a=90 h=7 pct ' ') value=(h=1) order=(0 to &v5max by 0.05)
offset=(1,1);
axis23 label=(a=90 h=7 pct ' ') value=(h=1) order=(0 to &v6max by 0.05)
offset=(1,1);
axis24 label=(h=1.2 pct ' ') value=(h=1);


symbol1 i=join w=1.5 line=1 v=dot color=black h=1;
symbol2 i=join w=1.5 line=2 v=dot  color=green h=1;
symbol3 i=join w=1.5 line=4 v=dot color=blue h=1;
symbol4 i=join w=1.5 line=20 v=dot color=red h=1;


/*
MSE_Dif_Impact_M1_M0
MSE_Impact_on_M0
MSE_Impact_on_T0

Bias_Dif_Impact_M1_M0
Bias_Impact_M0
Bias_Impact_T0
*/
proc gplot data=&datain;
plot MSE_Dif_Impact_M1_M0*proportion_TM=methods/name=&name1 vaxis= axis11
haxis=axis14 LEGEND=lengend1;
where Proportion_T eq &p1 and Proportion_M ne &p1;
note
f='Times New Roman' m=(8,40)pct h=1.2 a=90 'MSE('
f=greek  h=2 't'
f=simplex h=3  m=(+1.1,-1.6) "^"
f=ITALIC m=(-0.8,-0.9) h=1 'T'
f=greek m=(+.0, -.2) h=.5 '1'
f=simplex m=(-.0, +.2)  h=1 'x'
f=ITALIC  h=1 'T'
f=greek m=(+.0,-.2) h=.5 '2'
f='Times New Roman' m=(-0.3, +0.5) h=1.2   ')';
note
f='Times New Roman' m=(46,2.5)pct h=1.2 'Prevalance of'
f=ITALIC m=(+0,-0) h=1.2  'T'
f=greek  m=(-0.3,-0.2) h=0.8  '1'
f=ITALIC m=(+0.3,+0.2) h=1.2  ',T'
f=greek  m=(-0.3,-0.2) h=0.8  "2";
FOOTNOTE  m=(44,0)pct f='Times New Roman' h=1.2  '(a):MSE of ACIE
Estimates';
```

```
run;


proc gplot data=&datain;
plot Bias_Dif_Impact_M1_M0*proportion_TM=methods/name=&name2 vaxis=
axis21 haxis=axis24 LEGEND=lengend1;
where Proportion_T eq &p1 and Proportion_M ne &p1;
note
f='Times New Roman' m=(8,40)pct h=1.2 a=90 'RB('
f=greek  h=2 't'
f=simplex h=3  m=(+1.1,-1.6) "^"
f=ITALIC m=(-0.8,-0.9) h=1 'T'
f=greek m=(+.0, -.2) h=.5 '1'
f=simplex m=(-.0, +.2)  h=1 'x'
f=ITALIC  h=1 'T'
f=greek m=(+.0,-.2) h=.5 '2'
f='Times New Roman' m=(-0.3, +0.5) h=1.2   ')';
note
f='Times New Roman' m=(46,2.5)pct h=1.2 'Prevalance of'
f=ITALIC m=(+0,-0) h=1.2  'T'
f=greek  m=(-0.3,-0.2) h=0.8  '1'
f=ITALIC m=(+0.3,+0.2) h=1.2  ',T'
f=greek  m=(-0.3,-0.2) h=0.8  "2";
FOOTNOTE  m=(44,0)pct f='Times New Roman' h=1.2  '(b):RB of ACIE
Estimates';
run;



proc gplot data=&datain;
plot MSE_Impact_on_M0*proportion_TM=methods/name=&name3 vaxis= axis12
haxis=axis14 LEGEND=lengend1;
where Proportion_T eq &p1 and Proportion_M ne &p1;
note
f='Times New Roman' m=(8,40)pct h=1.2 a=90 'MSE('
f=greek  h=2 't'
f=simplex h=3  m=(+1.1,-1.6) "^"
f=ITALIC m=(-0.8,-0.9) h=1 'T'
f=greek m=(+0, -.2) h=.5 '1'
f='Times New Roman' m=(-0.3, +0.5) h=1.2   ')';
note
f='Times New Roman' m=(46,2.5)pct h=1.2 'Prevalance of'
f=ITALIC m=(+0,-0) h=1.2  'T'
f=greek  m=(-0.3,-0) h=0.8  '1'
f=ITALIC m=(+0.3,+0) h=1.2  ',T'
f=greek  m=(-0.3,-0) h=0.8  "2";
FOOTNOTE  m=(44,0)pct f='Times New Roman' h=1.2  '(c):MSE of ACME
Estimates for ' f=ITALIC m=(+0,-0) h=1.2  'T' m=(-0,-0) f=greek h=0.8
'1';

run;
```

```
proc gplot data=&datain;
plot Bias_Impact_M0*proportion_TM=methods/name=&name4 vaxis= axis22
haxis=axis24 LEGEND=lengend1;
where Proportion_T eq &p1 and Proportion_M ne &p1;
note
f='Times New Roman' m=(8,40)pct h=1.2 a=90 'RB('
f=greek  h=2 't'
f=simplex h=3  m=(+1.1,-1.6) "^"
f=ITALIC m=(-0.8,-0.9) h=1 'T'
f=greek m=(+0, -.2) h=.5 '1'
f='Times New Roman' m=(-0.3, +0.5) h=1.2   ')';
note
f='Times New Roman' m=(46,2.5)pct h=1.2 'Prevalance of'
f=ITALIC m=(+0,-0) h=1.2  'T'
f=greek  m=(-0.3,-0) h=0.8  '1'
f=ITALIC m=(+0.3,+0) h=1.2  ',T'
f=greek  m=(-0.3,-0) h=0.8  "2";
FOOTNOTE  m=(44,0)pct f='Times New Roman' h=1.2  '(d):RB of ACME
Estimates for ' f=ITALIC m=(+0,-0) h=1.2  'T' m=(-0,-0) f=greek h=0.8
'1';
run;

proc gplot data=&datain;
plot MSE_Impact_on_T0*proportion_TM=methods/name=&name5 vaxis= axis13
haxis=axis14 LEGEND=lengend1;
where Proportion_T eq &p1 and Proportion_M ne &p1;
note
f='Times New Roman' m=(8,40)pct h=1.2 a=90 'MSE('
f=greek  h=2 't'
f=simplex h=3  m=(+1.1,-1.6) "^"
f=ITALIC m=(-0.8,-0.9) h=1 'T'
f=greek m=(+0, -.2) h=.5 '2'
f='Times New Roman' m=(-0.3, +0.5) h=1.2   ')';
note
f='Times New Roman' m=(46,2.5)pct h=1.2 'Prevalance of'
f=ITALIC m=(+0,-0) h=1.2  'T'
f=greek  m=(-0.3,-0) h=0.8  '1'
f=ITALIC m=(+0.3,+0) h=1.2  ',T'
f=greek  m=(-0.3,-0) h=0.8  "2";
FOOTNOTE  m=(44,0)pct f='Times New Roman' h=1.2  '(e):MSE of ACME
Estimates for ' f=ITALIC m=(+0,-0) h=1.2  'T' m=(-0,-0) f=greek h=0.8
'2';

run;


proc gplot data=&datain;
plot Bias_Impact_T0*proportion_TM=methods/name=&name6 vaxis= axis23
haxis=axis24 LEGEND=lengend1;
where Proportion_T eq &p1 and Proportion_M ne &p1;
note
```

```
f='Times New Roman' m=(8,40)pct h=1.2 a=90 'RB('
f=greek  h=2 't'
f=simplex h=3  m=(+1.1,-1.6) "^"
f=ITALIC m=(-0.8,-0.9) h=1 'T'
f=greek m=(+0, -.2) h=.5 '2'
f='Times New Roman' m=(-0.3, +0.5) h=1.2   ')';
note
f='Times New Roman' m=(46,2.5)pct h=1.2 'Prevalance of'
f=ITALIC m=(+0,-0) h=1.2  'T'
f=greek  m=(-0.3,-0) h=0.8  '1'
f=ITALIC m=(+0.3,+0) h=1.2  ',T'
f=greek  m=(-0.3,-0) h=0.8  "2";
FOOTNOTE  m=(44,0)pct f='Times New Roman' h=1.2  '(f):RB of ACME
Estimates for ' f=ITALIC m=(+0,-0) h=1.2  'T' m=(-0,-0) f=greek h=0.8
'2';
run;
%mend;


/*Special chacters for label solution*/



%plotchap3_eq(simu.chap3_1,'plot_11','plot_12','plot_13','plot_14','plot_
15','plot 16');
%plotchap3_ne(simu.chap3_1,0.2,3,2,2,0.3,0.3,0.3,'plot_21','plot_22','plo
t_23','plot_24','plot_25','plot_26');
%plotchap3_ne(simu.chap3_1,0.3,3,2,2,0.3,0.3,0.3,'plot_31','plot_32','plo
t_33','plot_34','plot_35','plot_36');
%plotchap3_ne(simu.chap3_1,0.4,3,2,2,0.3,0.3,0.3,'plot_41','plot_42','plo
t_43','plot_44','plot_45','plot_46');
%plotchap3_ne(simu.chap3_1,0.5,3,2,2,0.3,0.3,0.3,'plot_51','plot_52','plo
t_53','plot_54','plot_55','plot_56');




%let gout=gout=work.gseg;
filename gsasfile 'C:\Users\Zirui Gu\Desktop\test\result\';
options nodate nonumber ;
goptions device=PNG300 gaccess=gsasfile xmax=8in ymax=7.5in
gsfmode=append  display
  rotate=landscape;


ODS RTF FILE='C:\Users\Zirui
Gu\Desktop\test\result\Chap3_Result_subtitle.RTF' BODYTITLE;


proc greplay nofs tc=work.tempcat igout=work.gseg &gout;
tdef t313r des='three left three right'
  1/llx=0 lly=67 ulx=0 uly=100 lrx=50 lry=67 urx=50 ury=100
```

```
  2/llx=50 lly=67 ulx=50 uly=100 lrx=100 lry=67 urx=100 ury=100
  3/llx=0 lly=34 ulx=0 uly=67 lrx=50 lry=34 urx=50 ury=67
  4/llx=50 lly=34 ulx=50 uly=67 lrx=100 lry=34 urx=100 ury=67
  5/llx=0 lly=0 ulx=0 uly=34 lrx=50 lry=0 urx=50 ury=34
  6/llx=50 lly=0 ulx=50 uly=34 lrx=100 lry=0 urx=100 ury=34;
  template t3l3r;
  treplay 1:plot_11 2:plot_12 3:plot_13 4:plot_14 5:plot_15 6:plot_16;
run;


proc greplay igout= GSEG nofs tc=work.tempcat;
template t3l3r;
treplay 1:plot_21 2:plot_22 3:plot_23 4:plot_24 5:plot_25 6:plot_26;
run;
quit;

proc greplay igout= GSEG nofs tc=work.tempcat;
template t3l3r;
treplay 1:plot_31 2:plot_32 3:plot_33 4:plot_34 5:plot_35 6:plot_36;
run;
quit;

proc greplay igout= GSEG nofs tc=work.tempcat;
template t3l3r;
treplay 1:plot_41 2:plot_42 3:plot_43 4:plot_44 5:plot_45 6:plot_46;
run;
quit;

proc greplay igout= GSEG nofs tc=work.tempcat;
template t3l3r;
treplay 1:plot_51 2:plot_52 3:plot_53 4:plot_54 5:plot_55 6:plot_56;
run;
quit;

ODS RTF CLOSE;
```

/*This Part Produce Figures for Chapter 4*/

```
options nonotes;

%macro
plotcaliper2(datain,model,height,ylabel,p11,p12,p21,p22,vaxis_range,var,l
abel,name1,name2,name3,name4);

Legend1 label=(height=1 position=top justify=center
 '')
 value=(h=2 'NNCVWR, m=1' 'NNCVWR, m=2' 'NNCVWR, m=3' 'NNCVWR, m=4'
'NNCVWR, m=5')
 across=1 down=5
 position = (top left inside)
 mode=protect
 offset = (3 pct);

Legend2 label=(height=1 position=top justify=center
 '')
 value=(h=2 'NNCV, m=1' 'NNCV, m=2' 'NNCV, m=3' 'NNCV, m=4' 'NNCV, m=5')
 across=1 down=5
 position = (top left inside)
 mode=protect
 offset = (3 pct);

axis11 label=(h=7 angle=90  "") value=(h=&height) order=(0 to 1 by 0.5)
offset=(1,1);
axis2 label=(h=2 f='Times New Roman' "w:Weight of Standard Deviation for
a Caliper Width")value=(h=1.5) order=(0 to 3 by 0.5);

symbol1 i=join w=2 line=1 v=none color=black h=6;
symbol2 i=join w=2 line=2 v=none  color=green h=6;
symbol3 i=join w=2 line=4 v=none color=blue h=6;
symbol4 i=join w=2 line=20 v=none color=red h=6;
symbol5 i=join w=2 line=41 v=none color=pink h=6;

title h=2 "Prevalances of 2 Treatments are &p11,&p12";
proc gplot data=&datain;
plot &var*Caliper=method/name=&name1 vaxis= axis11 haxis=axis2
Legend=Legend1;
where model=&model and replacement eq 'Replacement' and proportion_T eq
&p11 and proportion_M eq &p12;
note
f='Times New Roman' m=(0,40)pct h=2 a=90 'RB ( '
f=greek  h=3.5 't'
f=simplex h=4.5  m=(+2.3,-2.6) "^"
f=ITALIC m=(-2,-1.5) h=1.5 'T'
f=greek m=(+.0, +.05) h=1 '1'
f=simplex m=(-.0, +.05)  h=1.5 'x'
f=ITALIC  h=1.5 m=(-.0, +.05) 'T'
f=greek m=(+.0,+.05) h=1 '2'
f='Times New Roman' m=(-0.5, +0.8) h=2   ')';
```

```
run;


title h=2 "Prevalances of 2 Treatments are &p11,&p12";
proc gplot data=&datain;
plot &var*Caliper=method/name=&name2 vaxis= axis11 haxis=axis2
Legend=Legend2;
where model=&model and replacement eq 'Without Replacement' and
proportion_T eq &p11 and proportion_M eq &p12;
note
f='Times New Roman' m=(0,40)pct h=2 a=90 'RB ( '
f=greek  h=3.5 't'
f=simplex h=4.5  m=(+2.3,-2.6) "^"
f=ITALIC m=(-2,-1.5) h=1.5 'T'
f=greek m=(+.0, +.05) h=1 '1'
f=simplex m=(-.0, +.05)  h=1.5 'x'
f=ITALIC  h=1.5 m=(-.0, +.05) 'T'
f=greek m=(+.0,+.05) h=1 '2'
f='Times New Roman' m=(-0.5, +0.8) h=2   ')';
run;


title h=2 "Prevalances of 2 Treatments are &p21,&p22";
proc gplot data=&datain;
plot &var*Caliper=method/name=&name3 vaxis= axis11 haxis=axis2
Legend=Legend1;
where model=&model and replacement eq 'Replacement' and proportion_T eq
&p21 and proportion_M eq &p22;
note
f='Times New Roman' m=(0,40)pct h=2 a=90 'RB ( '
f=greek  h=3.5 't'
f=simplex h=4.5  m=(+2.3,-2.6) "^"
f=ITALIC m=(-2,-1.5) h=1.5 'T'
f=greek m=(+.0, +.05) h=1 '1'
f=simplex m=(-.0, +.05)  h=1.5 'x'
f=ITALIC  h=1.5 m=(-.0, +.05) 'T'
f=greek m=(+.0,+.05) h=1 '2'
f='Times New Roman' m=(-0.5, +0.8) h=2   ')';
run;


title h=2 "Prevalances of 2 Treatments are &p21,&p22";
proc gplot data=&datain;
plot &var*Caliper=method/name=&name4 vaxis= axis11 haxis=axis2
Legend=Legend2;
where model=&model and replacement eq 'Without Replacement' and
proportion_T eq &p21 and proportion_M eq &p22;
note
f='Times New Roman' m=(0,40)pct h=2 a=90 'RB ( '
f=greek  h=3.5 't'
f=simplex h=4.5  m=(+2.3,-2.6) "^"
f=ITALIC m=(-2,-1.5) h=1.5 'T'
```

```
f=greek m=(+.0, +.05) h=1 '1'
f=simplex m=(-.0, +.05)  h=1.5 'x'
f=ITALIC  h=1.5 m=(-.0, +.05) 'T'
f=greek m=(+.0,+.05) h=1 '2'
f='Times New Roman' m=(-0.5, +0.8) h=2   ')';
run;
%mend;

%macro plotcaliperpack(data,model,ylabel,measurevar,height,filename);
%plotcaliper2(&data,&model,1,&ylabel,0.2,0.2,0.3,0.3,&height,&measurevar,
&label1,'NAG22__1','NAG22__2','NAG33__1','NAG33__2');
%plotcaliper2(&data,&model,1,&ylabel,0.4,0.4,0.5,0.5,&height,&measurevar,
&label1,'NAG44__1','NAG44__2','NAG55__1','NAG55__2');
%plotcaliper2(&data,&model,1,&ylabel,0.6,0.6,0.7,0.7,&height,&measurevar,
&label1,'NAG66__1','NAG66__2','NAG77__1','NAG77__2');
%plotcaliper2(&data,&model,1,&ylabel,0.8,0.8,0.2,0.3,&height,&measurevar,
&label1,'NAG88__1','NAG88__2','NAG23__1','NAG23__2');
%plotcaliper2(&data,&model,1,&ylabel,0.2,0.4,0.2,0.5,&height,&measurevar,
&label1,'NAG24__1','NAG24__2','NAG25__1','NAG25__2');
%plotcaliper2(&data,&model,1,&ylabel,0.2,0.6,0.2,0.7,&height,&measurevar,
&label1,'NAG26__1','NAG26__2','NAG27__1','NAG27__2');
%plotcaliper2(&data,&model,1,&ylabel,0.2,0.8,0.3,0.2,&height,&measurevar,
&label1,'NAG28__1','NAG28__2','NAG32__1','NAG32__2');
%plotcaliper2(&data,&model,1,&ylabel,0.3,0.4,0.3,0.5,&height,&measurevar,
&label1,'NAG34  1','NAG34  2','NAG35  1','NAG35  2');
%plotcaliper2(&data,&model,1,&ylabel,0.3,0.6,0.3,0.7,&height,&measurevar,
&label1,'NAG36__1','NAG36__2','NAG37__1','NAG37__2');
%plotcaliper2(&data,&model,1,&ylabel,0.3,0.8,0.4,0.2,&height,&measurevar,
&label1,'NAG38__1','NAG38__2','NAG42__1','NAG42__2');
%plotcaliper2(&data,&model,1,&ylabel,0.4,0.3,0.4,0.5,&height,&measurevar,
&label1,'NAG43__1','NAG43__2','NAG45__1','NAG45__2');
%plotcaliper2(&data,&model,1,&ylabel,0.4,0.6,0.4,0.7,&height,&measurevar,
&label1,'NAG46__1','NAG46__2','NAG47__1','NAG47__2');
%plotcaliper2(&data,&model,1,&ylabel,0.4,0.8,0.5,0.2,&height,&measurevar,
&label1,'NAG48__1','NAG48__2','NAG52__1','NAG52__2');
%plotcaliper2(&data,&model,1,&ylabel,0.5,0.3,0.5,0.4,&height,&measurevar,
&label1,'NAG53__1','NAG53__2','NAG54__1','NAG54__2');
%plotcaliper2(&data,&model,1,&ylabel,0.5,0.6,0.5,0.7,&height,&measurevar,
&label1,'NAG56__1','NAG56__2','NAG57__1','NAG57__2');
%plotcaliper2(&data,&model,1,&ylabel,0.5,0.8,0.5,0.8,&height,&measurevar,
&label1,'NAG58__1','NAG58__2','NAG59__1','NAG59__2');
ODS RTF FILE=&filename;

%let gout=gout=work.gseg;
options nodate nonumber ;
goptions device=PNG300 gaccess=gsasfile xmax=8in ymax=9in gsfmode=append
display
  rotate=landscape;

proc greplay nofs tc=work.tempcat igout=work.gseg &gout;
tdef t4l4r des='Four left Four right'
  1/llx=12.5 lly=80 ulx=12.5  uly=100 lrx=47.5  lry=80 urx=47.5 ury=100
```

```
  2/llx=52.5 lly=80 ulx=52.5  uly=100 lrx=87.5  lry=80 urx=87.5 ury=100
  3/llx=12.5 lly=55 ulx=12.5  uly=75  lrx=47.5  lry=55 urx=47.5 ury=75
  4/llx=52.5 lly=55 ulx=52.5  uly=75  lrx=87.5  lry=55 urx=87.5 ury=75
  5/llx=12.5 lly=30 ulx=12.5  uly=50  lrx=47.5  lry=30 urx=47.5 ury=50
  6/llx=52.5 lly=30 ulx=52.5  uly=50  lrx=87.5  lry=30 urx=87.5 ury=50
  7/llx=12.5 lly=5  ulx=12.5  uly=25  lrx=47.5  lry=5  urx=47.5 ury=25
  8/llx=52.5 lly=5  ulx=52.5  uly=25  lrx=87.5  lry=5  urx=87.5 ury=25;
  template t4l4r;
  treplay 1:NAG22__1 2:NAG22__2 3:NAG88__1 4:NAG88__2 5:NAG33__1
6:NAG33__2 7:NAG77__1 8:NAG77__2;
run;


options nodate nonumber ;
goptions device=PNG300 gaccess=gsasfile xmax=6in ymax=6.75in
gsfmode=append  display
  rotate=landscape;
%let gout=gout=work.gseg;
proc greplay nofs tc=work.tempcat igout=work.gseg &gout;
tdef t3l3r des='three left three right'
  1/llx=0    lly=73.26 ulx=0     uly=100     lrx=46.8  lry=73.26
urx=46.8 ury=100
  2/llx=53.2 lly=73.26 ulx=53.2  uly=100     lrx=100   lry=73.26
urx=100  ury=100
  3/llx=0    lly=40.12 ulx=0     uly=66.86   lrx=46.8  lry=40.12
urx=46.8 ury=66.86
  4/llx=53.2 lly=40.12 ulx=53.2  uly=66.86   lrx=100   lry=40.12
urx=100  ury=66.86
  5/llx=0    lly=8.98  ulx=0     uly=35.72   lrx=46.8  lry=8.98
urx=46.8 ury=35.72
  6/llx=53.2 lly=8.98  ulx=53.2  uly=35.72   lrx=100   lry=8.98
urx=100  ury=35.72;
  template t3l3r;
  treplay 1:NAG44__1 2:NAG44__2 3:NAG66__1 4:NAG66__2 5:NAG55__1
6:NAG55__2;
run;


goptions device=PNG300 gaccess=gsasfile xmax=8in ymax=9in gsfmode=append
display
  rotate=landscape;

proc greplay nofs tc=work.tempcat igout=work.gseg &gout;
tdef t4l4r des='Four left Four right'
  1/llx=12.5 lly=80 ulx=12.5  uly=100 lrx=47.5  lry=80 urx=47.5 ury=100
  2/llx=52.5 lly=80 ulx=52.5  uly=100 lrx=87.5  lry=80 urx=87.5 ury=100
  3/llx=12.5 lly=55 ulx=12.5  uly=75  lrx=47.5  lry=55 urx=47.5 ury=75
  4/llx=52.5 lly=55 ulx=52.5  uly=75  lrx=87.5  lry=55 urx=87.5 ury=75
  5/llx=12.5 lly=30 ulx=12.5  uly=50  lrx=47.5  lry=30 urx=47.5 ury=50
  6/llx=52.5 lly=30 ulx=52.5  uly=50  lrx=87.5  lry=30 urx=87.5 ury=50
  7/llx=12.5 lly=5  ulx=12.5  uly=25  lrx=47.5  lry=5  urx=47.5 ury=25
  8/llx=52.5 lly=5  ulx=52.5  uly=25  lrx=87.5  lry=5  urx=87.5 ury=25;
```

```
  template t4l4r;
  treplay 1:NAG23__1 2:NAG23__2 3:NAG32__1 4:NAG32__2 5:NAG24__1
6:NAG24__2 7:NAG42__1 8:NAG42__2;
run;



proc greplay igout= GSEG nofs tc=work.tempcat;
template t4l4r;
treplay 1:NAG25__1 2:NAG25__2 3:NAG52__1 4:NAG52__2 5:NAG34__1 6:NAG34__2
7:NAG43__1 8:NAG43__2;
run;
quit;

proc greplay igout= GSEG nofs tc=work.tempcat;
template t4l4r;
treplay 1:NAG26__1 2:NAG26__2 3:NAG27__1 4:NAG27__2 5:NAG28__1 6:NAG28__2
7:NAG36__1 8:NAG36__2;
run;
quit;

proc greplay igout= GSEG nofs tc=work.tempcat;
template t4l4r;
treplay 1:NAG35__1 2:NAG35__2 3:NAG53__1 4:NAG53__2 5:NAG45__1 6:NAG45__2
7:NAG54__1 8:NAG54__2;
run;
quit;

proc greplay igout= GSEG nofs tc=work.tempcat;
template t4l4r;
treplay 1:NAG37__1 2:NAG37__2 3:NAG38__1 4:NAG38__2 5:NAG46__1 6:NAG46__2
7:NAG47__1 8:NAG47__2;
run;
quit;

proc greplay igout= GSEG nofs tc=work.tempcat;
template t4l4r;
treplay 1:NAG48__1 2:NAG48__2 3:NAG56__1 4:NAG56__2 5:NAG57__1 6:NAG57__2
7:NAG58__1 8:NAG58__2;
run;
quit;



ODS RTF CLOSE;
%mend;

libname simu  'C:\Users\Zirui Gu\Desktop\test\result\';
```

```
/*
data simu.Solu_test4;set simu.Solu_test2;
*length proportion_TM $50. method $50.;
*proportion_TM=strip(PROPORTION_T)||','||Strip(PROPORTION_M);
method=note2;
if Caliper ne . then method='NN Caliper
'||strip(replacement)||',m='||strip(m)||','||strip(model);
run;


proc sort data=simu.Solu_test4;
by  proportion_TM method caliper;
run;
*/



%plotcaliperpack(simu.Solu_test4,'ANOVA','RB ( ',Bias_Dif_Impact_M1_M0,1,
'C:\Users\Zirui Gu\Desktop\test\result\Chap4_RB_Interaction1.RTF');
```

/*This part produce tables for chapter 4*/

```
%macro calitable(datain,dataout);

data calionly;set &datain;
where model eq 'ANOVA' and caliper ne . and caliper le 3;
length method3 $200.;
method3=strip(replacement)||strip(proportion_tm);
run;
proc sort data=calionly;
by method3 Bias_Dif_Impact_M1_M0;
run;


data calionly;set calionly;
by method3;
if first.method3 then output;
run;

proc sql noprint;
create table temp1 as select proportion_TM, replacement,caliper,
m,Bias_Dif_Impact_M1_M0
from calionly
order by proportion_tm,replacement,m;
quit;
data temp1(keep=proportion_TM caliper m replacement
Bias_Dif_Impact_M1_M0);set calionly;
run;

data temp1_1;set temp1;
```

```
where replacement eq 'Replacement';
run;

data temp1_2;set temp1;
rename Bias_Dif_Impact_M1_M0 = Bias_Dif_Impact_M1_M02;
rename Caliper = Caliper2;
rename m=m2;
where replacement ne 'Replacement';
drop replacement;
run;

proc sql noprint;
create table newtemp as select * from temp1_1 a ,temp1_2 b
where a.proportion_tm =b.proportion_tm;
quit;

data newtemp2;set newtemp;
length MSE_compare $20. cali_compare $20.;
MSE_compare =
substr(strip(Bias_Dif_Impact_M1_M0),1,4)||'/'||substr(strip(Bias_Dif_Impa
ct_M1_M02),1,4);
cali_compare= strip(Caliper)||'/'||strip(Caliper2);
m_compare=strip(m)||'/'||strip(m2);
drop Bias_Dif_Impact_M1_M0 Bias_Dif_Impact_M1_M02 caliper caliper2
replacement m m2;
length methods $200.;
methods='NNCVWR/NNCV';
*if proportion_tm in (&p1,&p2) then order=1;
*if proportion_tm in (&p3,&p4) then order=2;
run;

proc sql noprint;
create table &dataout as select Proportion_tm,
methods,m_compare,cali_compare,MSE_compare
from newtemp2
order by Proportion_tm ;
quit;

proc print data=&dataout noobs;
run;
%mend;

options nonotes;
%calitable(simu.Solu_test4,simu.Opt_cali_MSE_ACIE);


options notes;
proc sort data= simu.chap4_3;
by proportion_tm MSE_Dif_Impact_M1_M0;
run;

data simu.chap4_3uni1;set simu.chap4_3;
```

```
by proportion_tm;
if first.proportion_tm then output;
run;



proc sort data= simu.chap4_3;
by proportion_tm MSE_Impact_on_M0;
run;

data simu.chap4_3uni2;set simu.chap4_3;
by proportion_tm;
if first.proportion_tm then output;
run;



proc sort data= simu.chap4_3;
by proportion_tm MSE_Impact_on_T0;
run;

data simu.chap4_3uni3;set simu.chap4_3;
by proportion_tm;
if first.proportion_tm then output;
run;


proc sql noprint;
create table simu.chap4_table410 as select a.proportion_tm, a.methods as
MSE_ACIE, b.methods as MSE_ACME1, c.methods as MSE_ACME2
from simu.chap4_3uni1 a, simu.chap4_3uni2 b, simu.chap4_3uni3 c
where a.proportion_tm eq b.proportion_tm eq c.proportion_tm
order by a.proportion_tm;
quit;


proc print data=simu.chap4_table410 noobs;
run;


proc sort data= simu.chap4_3;
by proportion_tm Bias_Impact_T0;
run;

data simu.chap4_3uni;set simu.chap4_3;
by proportion_tm;
if first.proportion_tm then output;
run;

proc freq data=simu.chap4_3uni;
tables methods;
run;
```

```
data check;set simu.chap4_3;
where methods in ('6.NNCV','3.NNWR');
run;


proc sort data= check;
by proportion_tm Bias_Dif_Impact_M1_M0;
run;

data check;set check;
by proportion_tm;
if first.proportion_tm then output;
run;

proc freq data=check;
tables methods;
run;
```

VITA

Zirui Gu is a US permanent resident born in China on 3/8/1986. He received bachelor of science degree in Educational Software Engineering in Capital Normal University in 2008. He obtained his master degree in statistics at department of mathematics and statistics in Mississippi State University. He served as a biostatistician in VCUHS from May, 2013 to September, 2016. He is currently a data scientist for the division of population health management in Siemens Healthineers.